Cognitive Psychology 101 (2018) 50–81

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Conditionals and inferential connections: A hypothetical inferential theory☆

Igor Douven[a,1], Shira Elqayam[b,*,1], Henrik Singmann[c,1], Janneke van Wijnbergen-Huitink[d]

[a] SND/CNRS/Sorbonne University, France
[b] School of Applied Social Sciences, De Montfort University, United Kingdom
[c] Institute of Psychology, University of Zürich, Switzerland
[d] Faculty of Philosophy, University of Nijmegen, Netherlands

ARTICLE INFO

Keywords:
Belief bias
Conditionals
Dual processing
Inferential semantics
Relevance
Satisficing

ABSTRACT

Intuition suggests that for a conditional to be evaluated as true, there must be some kind of connection between its component clauses. In this paper, we formulate and test a new psychological theory to account for this intuition. We combined previous semantic and psychological theorizing to propose that the key to the intuition is a relevance-driven, satisficing-bounded inferential connection between antecedent and consequent. To test our theory, we created a novel experimental paradigm in which participants were presented with a soritical series of objects, notably colored patches (Experiments 1 and 4) and spheres (Experiment 2), or both (Experiment 3), and were asked to evaluate related conditionals embodying non-causal inferential connections (such as "If patch number 5 is blue, then so is patch number 4"). All four experiments displayed a unique response pattern, in which (largely determinate) responses were sensitive to parameters determining inference strength, as well as to consequent position in the series, in a way analogous to belief bias. Experiment 3 showed that this guaranteed relevance can be suppressed, with participants reverting to the defective conditional. Experiment 4 showed that this pattern can be partly explained by a measure of inference strength. This pattern supports our theory's "principle of relevant inference" and "principle of bounded inference," highlighting the dual processing characteristics of the inferential connection.

1. Introduction

Conditionals are sentences of the form "If φ, [then] ψ," with φ called "the antecedent" and ψ, "the consequent."[2] The functions of conditionals are many and varied. For instance, we use "if" when we want to think hypothetically about non-actual possibilities (Evans & Over, 2004); we use "if" to express causal relations (Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007) or probabilistic relations (Evans, Handley, & Over, 2003; Oberauer & Wilhelm,

☆ The supplemental materials as well as all data and all analysis scripts are available at: https://osf.io/3uajq/.

* Corresponding author.
E-mail addresses: selqayam@dmu.ac.uk, selqayam@hotmail.com (S. Elqayam).

[1] The first three authors contributed equally to the paper and are listed alphabetically.

[2] Unless stated otherwise, "conditional" refers to *indicative* conditionals. Indicative conditionals are conditionals whose antecedent is in the indicative mood. They are standardly contrasted with subjunctive conditionals, whose antecedent is in the subjunctive mood. In this paper, we are only concerned with indicative conditionals, although in Experiment 2 we address a possible subjunctive reading of some of our materials.

https://doi.org/10.1016/j.cogpsych.2017.09.002
Accepted 25 September 2017
Available online 09 January 2018

2003); and we use "if" to express pragmatic functions such as inducement or advice (Beller, Bender, & Kuhnmünch, 2005; Evans, Neilens, Handley, & Over, 2008; Fillenbaum, 1976, 1986), and to generate novel normative rules (Elqayam, Thompson, Wilkinson, Evans, & Over, 2015). It is no wonder, then, that the study of conditionals has engaged psychologists and philosophers alike. Without a good theory of conditionals, we have no hope of understanding human reasoning or decision making. In this paper, we address what is arguably *the* most central question in the study of conditionals, to wit, how the antecedent connects to the consequent. Specifically, we will be concerned with how people's judgments of the truth values of conditionals vary as a function of the link between antecedent and consequent.

Intuitively, when we state a conditional, we expect that the antecedent would be relevant to the consequent. For example, there is something odd about the following conditionals:

(1) a. If Isaac Newton preferred apples over oranges, then he got his best ideas while walking.
b. If Winston Churchill did not sleep the night before D-Day, then he considered a career as a sculptor early on in life.

These conditionals appear odd in that the truth of their antecedent seems irrelevant to their consequent. There is no intelligible notion of dependency in which Newton's having gotten his best ideas while walking could be said to have depended on whether he preferred apples over oranges, and similarly for (1b): whether the young Churchill considered a career as a sculptor can hardly have depended on how he slept the night before D-Day.

Although both dependency and relevance have played key roles in psychological theories of conditionals (e.g., Evans & Over, 2004; Oaksford & Chater, 1994), there has been no systematic effort to explore the psychological mechanisms that make sentences such as

(2) If global warming continues, then parts of England will be flooded.

seem plausible, where sentences such as (1a) and (1b) are not. In this paper, we will formulate and support a psychological account of the relation between antecedent and consequent, to explain why conditionals like (1a) and (1b) strike us as odd, and what this tells us about the psychological mechanisms underlying our understanding of conditionals. To do this, we propose a new theory that combines insights from two main theoretical accounts: the philosophical account of inferentialism, and the psychological theory of hypothetical thinking.

Where virtually all semantics of conditionals define the truth values of conditionals as functions of the truth values of the conditionals' antecedents and consequents (whether in the actual world or also in other possible worlds), inferentialism is the only semantics that makes the existence of an inferential connection between antecedent and consequent a requirement for the truth of a conditional. Inferentialism, in other words, builds the requirement of a connection into the *meaning* of the word "if," thereby straightforwardly accounting for the felt oddness of conditionals such as (1a) and (1b). It is not that these conditionals appear odd because they are semantically defective. A sentence can appear perfectly fine while still being semantically defective—the world may simply fail to cooperate. According to inferentialism, (1a) and (1b) appear odd because they are semantically defective for a reason that could easily have been avoided: it is (typically) under our control to compose conditionals whose component parts stand in an inferential relationship to one another.

Hypothetical thinking theory (Evans, 2006, 2007a, 2010) is a dual-process theory positing two types of processes: fast, resource-frugal, and intuitive processes, and slow, analytic processes. The former generate the most relevant, single mental representation; the latter can then intervene to revise or even reject the initial representation, but this is a lazy, bounded process, meaning that the initial representation tends to be adopted unless compelling reasons for revision exist.

Our blend of inferentialism with dual processing, and more specifically with hypothetical thinking theory, allows us to hypothesize that the connection between antecedent and consequent is an inferential one, governed by relevance and bounded by satisficing. We state this theory in detail in a separate section, and present evidence in its favor from four experiments. But we begin by reviewing the main extant psychological accounts of conditionals, with some reference along the way to relevant philosophical accounts as well.

2. Theories of conditionals

2.1. Mental model theory and the material conditional account

Philosophical theorizing about the semantics for conditionals has long been dominated by the material conditional account, as advocated by, among others, Grice (1989), Jackson (1979), and Lewis (1976). According to this account, the truth conditions of a conditional are those of the corresponding material conditional: "If φ, ψ" is false if φ is true and ψ is false, and it is true in all other cases. Although the material conditional has several advantages, it has also been criticized for sanctioning a number of counter-intuitive inferences. Most famously, it gives rise to the so-called paradoxes of the material conditional: it validates the intuitively invalid inference of "If φ, ψ" from not-φ (e.g., the inference of "If Bill Gates went bankrupt, he is a billionaire" from "Bill Gates did not go bankrupt"), as well as the intuitively equally invalid inference of "If φ, ψ" from ψ (e.g., the inference of "If Bill Gates went bankrupt, he is a billionaire" from "Bill Gates is a billionaire"). It is fair to say that this account is no longer considered as the received doctrine among philosophers working on conditionals.

In psychology, the state of the art is similar, in that few psychological theories formulated past the turn of the century take the material conditional account as their starting point. The one exception is mental model theory, in which the basic (i.e., abstract)

conditional corresponds to the material conditional (Johnson-Laird & Byrne, 2002). However, the theory is supplemented by semantic and pragmatic modulations, so the material conditional only applies to a limited set of conditionals. We note that mental model theory has recently been radically revised (Johnson-Laird, Khemlani, & Goodwin, 2015), rejecting the paradoxes of the material conditional, although the new theory still needs more fleshing out (Baratgin et al., 2015).

2.2. The Ramsey test and the Equation

Except for theories belonging to the material conditional family, almost all contemporary theories of conditionals, both in psychology and in philosophy, build on the celebrated Ramsey test (Ramsey, 1929/1990). Supported by much psychological evidence (Evans & Over, 2004), the Ramsey test posits that we determine whether to accept a given conditional by hypothetically adding its antecedent to our stock of beliefs, making minimal changes (if necessary) to preserve consistency, and from the resulting (hypothetical) perspective judging the acceptability of the conditional's consequent. So, to evaluate (2), we hypothetically suppose that global warming continues, and evaluate under this supposition the acceptability that parts of England will be flooded.

Much theorizing constructed around the Ramsey test also subscribes to the Equation, which is suggested by the same footnote in (Ramsey, 1929/1990) that presents the Ramsey test. According to the Equation, the probability of a conditional, $\Pr(\text{If } \varphi, \psi)$, corresponds to the conditional probability $\Pr(\psi \mid \varphi)$. For example, the probability that if global warming continues, parts of England will be flooded, is the probability of parts of England being flooded given that global warming continues.

The one exception is Stalnaker's (1968) possible worlds semantics, which was inspired by the Ramsey test but does not commit to the Equation. According to Stalnaker, a conditional is true (false) if its consequent is true (false) in the closest possible world in which its antecedent is true—provided there is a world in which its antecedent is true; otherwise it is vacuously true. We are not aware of any psychological theory explicitly committed to Stalnaker's semantics, although it is one possible interpretation of the psychological suppositional conditional, which we describe in a separate section.

2.3. New paradigm and the equation

The Ramsey test and the Equation are both cornerstones of what has been dubbed "the New Paradigm" in psychology of reasoning (Elqayam, 2017; Elqayam & Over, 2013; Manktelow, Over, & Elqayam, 2011; Over, 2011). The traditional paradigm in psychology of reasoning focused on binary truth values (e.g., Johnson-Laird & Byrne, 2002). Its underlying semantic theory was, fittingly, the binary material conditional. One of the hallmarks of the New Paradigm is that this binary approach is replaced with the more psychological focus on uncertainty and subjective degrees of belief. In other words, reasoning in the New Paradigm is seen as Bayesian, at least to some extent (Douven, 2016a; Elqayam & Evans, 2013).

Within the Equation-oriented camp in philosophy, there is a further distinction between non-propositionalism, according to which conditionals do not express propositions and never have a truth value, and the three-value view. According to the latter, a conditional is true if it has both a true antecedent and a true consequent; false if it has a true antecedent and a false consequent; and neither true nor false ("void," "indeterminate") when its antecedent is false (Bennett, 2003; de Finetti, 1995). We are not aware of any psychological theory fully committed to non-propositionalism; the three-value view is one of the possible semantic interpretations underlying the psychological suppositional conditional.

2.4. Inferentialism

According to all of the truth-conditional semantics discussed so far, a sufficient condition for the truth of a conditional is that the conditional's antecedent and consequent are both true, no matter how internally unrelated these are. Thus, if Newton preferred apples over oranges and he got his best ideas while walking, then (1a) is true according to those semantics. In this respect, these semantics contrast sharply with the final semantics to be reviewed here, to wit, inferentialism, which holds that, for a conditional to be true, we should be able to *infer* its consequent from its antecedent (e.g., in philosophy, Barwise & Perry, 1983; Kratzer, 1986; Mill, 1843/1872; Ramsey, 1929/1990; Récanati, 2000; and in psychology, Braine, 1978; Braine & O'Brien, 1991).

As explained in the introduction, inferentialism makes it straightforward to account for our intuitions concerning (1a) and (1b). Nevertheless, the inferentialist approach to the semantics of conditionals has never enjoyed wide popularity, chiefly because critics have had no difficulty pointing at conditionals that are pre-theoretically true, yet whose consequent is seemingly *not* inferable from their antecedent. Consider, for instance,

(3) If Betty misses her bus, she will be late for the movies.

It is easy enough to imagine circumstances under which we would regard this conditional as true, even though we can never rule out that, through some freak accident, Betty makes it to the cinema in time even if she misses her bus. However, this objection has force only if we interpret "inference" as meaning *deductive* inference. Krzyżanowska, Wenmackers, and Douven (2014) proposed a version of inferentialism based on a notion of inference that goes beyond deduction: the argument from antecedent to consequent may contain not only deductive but also abductive and inductive inferential steps, where (roughly) abductive inference is inference based on explanatory considerations and inductive inference is inference based on statistical grounds. On this proposal,

(4) If Wilma and Fred are going to the gym together, they have settled their dispute.

may be true because, given relevant background knowledge, Wilma and Fred having settled their dispute is the best explanation of their going to the gym together. In the same way,

(5) If Barney works hard, he will pass the exam with flying colors.

may be true because Barney is a very bright student and typically when such students work hard for an exam, they pass it with flying colors.

For our purposes, this proposal may be summarized as stating that a conditional is true if and only if there exists a *strong enough* argument leading from its antecedent plus background knowledge to its consequent. What counts as strong enough may be subject to cognitive and contextual variations and is beyond the scope of the present work; for now, we will stick with a Simonian notion of satisficing (Simon, 1982), on which we will elaborate later. The idea that arguments need not be deductively valid and that informal arguments can still be judged for argument strength is fully compatible with the New Paradigm view of informal argumentation and its significance (e.g., Hahn & Oaksford, 2007; Mercier & Sperber, 2011), although Krzyżanowska et al. (2014) do not commit themselves to a Bayesian framework.[3]

3. The suppositional conditional and the defective truth table: The empirical evidence

As mentioned in the previous section, much relevant psychological work on conditionals has been carried out within the framework of the New Paradigm. In this section, we review the empirical evidence for the Ramsey test, the Equation, and the psychological theories that accommodate them. Many contemporary psychological theories accept the Ramsey test and the Equation as their starting point (see Evans & Over, 2004; Oaksford & Chater, 2007, for reviews). Evans and Over's (2004) psychological theory of the suppositional conditional is representative of this approach. Their proposal leaves the precise nature of the computational-level theory open: Evans and Over suggested that psychological findings tell decisively against the material conditional account, but can fit either Stalnaker's semantics or three-value theories (Baratgin, Over, & Politzer, 2013; Gilio & Over, 2012; Politzer, Over, & Baratgin, 2010; although we note that van Wijnbergen-Huitink, Elqayam, & Over, 2015, were later able to obtain direct empirical evidence for the Equation and against Stalnaker's account). They do not discuss inferentialism, but one could argue that position offers another way of fleshing out the thought that evaluating a conditional crucially relies on suppositional thinking. We take this up again in the next section.

Over the last decade or so, plenty of empirical support for the Equation has accumulated in the psychological literature, much of it coming from the probabilistic truth table task. Participants in this task are presented with a conditional and asked to estimate its probability. They are also given a probability distribution on the four truth table combinations (TT, TF, FT, and FF), or are asked to generate one themselves in a separate task (Evans et al., 2003; Oberauer & Wilhelm, 2003; Over et al., 2007). Almost invariably, the probability estimates of conditionals strongly correlate with the corresponding conditional probabilities computed on the basis of the truth table cases—which is what one expects to find if the Equation holds true (Douven & Verbrugge, 2010, 2013; Evans & Over, 2004; Fugard, Pfeifer, Mayerhofer, & Kleiter, 2011; Gauffroy and Barrioullet, 2009; Oaksford & Chater, 2003, 2007; Oberauer, Weidenfeld, & Fischer, 2007; Over, Douven, & Verbrugge, 2013; Over & Evans, 2003; Pfeifer & Kleiter, 2010; Politzer et al., 2010).

Of special interest to our context is the "defective" (or de Finetti) truth table (see Evans & Over, 2004, for a review; also Over & Baratgin, 2017; Over & Cruz, 2017). In this task, participants are asked to evaluate the four truth table combinations of conditional sentences. The idea goes back to Wason (1966), the founder of modern reasoning research, who suggested that reasoners regard the false antecedent cases as irrelevant. The defective truth table is the one with the pattern TF##; that is to say, TT (the case in which both the antecedent and consequent are true) is evaluated as true, TF as false, and both FT and FF as indeterminate. This pattern is prevalent when participants are presented with arbitrary, abstract conditionals such as "If there is a King on one side of the card, then there is a 3 on the other side." It can also be identified with the probabilistic truth table task, and it remains reliable when participants are asked to place bets rather than assign truth conditions (Baratgin et al., 2013; Politzer et al., 2010). The defective truth table is related to cognitive proficiency: the pattern becomes more prevalent with age (Barrouillet, Gauffroy, & Lecas, 2008); in adults, it correlates with general cognitive ability (Evans, Handley, Neilens, & Over, 2010); and it becomes more dominant as participants accrue practice (Fugard et al., 2011).

The probabilistic truth table task (Over et al., 2007) provides some analogous findings for thematic materials, although the comparison is not entirely straightforward. In this task, participants are given thematic conditionals—usually causal or diagnostic conditionals—and asked to evaluate their probability; in a separate truth table task, they are provided with each of the truth table cases and asked to evaluate their probabilities so that they sum to 100 percent. Typically, conditional probability computed based on evaluations of the TT and TF cases is found to be the single strongest predictor of estimates of the probability of the conditional, whereas measures based on the FT and FF cases—most importantly, the Δp rule, which measures the difference between $\Pr(\psi \mid \varphi)$ and

[3] See Douven (2016b, chap. 2), for a rebuttal of potential objections to inferentialism. For some first empirical results in support of an inferentialist semantics, see Vidal and Baratgin (2017). Cruz, Over, Oaksford, and Baratgin (2016) suggest that the requirement of an inferential connection between antecedent and consequent may be best accounted for in pragmatic terms. See Krzyżanowska, Collins, and Hahn (2017) for some evidence against this suggestion. In Douven, Elqayam, Singmann, and van Wijnbergen-Huitink (2017), we give some reasons for holding that inferentialist intuitions are best explained as emanating from the semantics, and not from the pragmatics, of conditionals. It is probably fair to say, though, that at the moment there is no conclusive argument in favor of either position. In any case, in the present paper we explicitly remain noncommittal on whether inferentialist intuitions have a semantic rather than a pragmatic origin. See also Douven and Krzyżanowska (in press) on some pitfalls of trying to distinguish experimentally semantic from pragmatic phenomena.

$\Pr(\psi \mid \text{not–}\varphi)$–are relatively poor predictors (see also Singmann, Klauer, & Over, 2014; but cf. Ohm & Thompson, 2006; Skovgaard-Olsen, Singmann, & Klauer, 2016).

4. Toward a new psychological theory of conditionals

The previous review should make it obvious that no single existing theory covers the full range of intuitions and psychological evidence. Specifically, inferentialism as proposed by Krzyżanowska et al. (2014) does not cover the defective truth table; and extant psychological theories of conditionals, such as the psychological suppositional conditional (Evans & Over, 2004), do not sufficiently cover inferentialist intuitions. Our aim in this paper is to construct and test a psychological theory which covers both. To do this, we will draw on features from inferentialism, combined with features taken from the psychological suppositional conditional, and its parent theory, hypothetical thinking theory (Evans, 2006, 2007a; Evans & Over, 2004). Thus, our theoretical account integrates algorithmic and computational aspects (in the sense of Marr, 1982)—processing and representational features on the one hand, and formal (semantic or pragmatic) features on the other hand, respectively. We call our theory "Hypothetical Inferential Theory," or HIT, for short.

Dual process theories of higher cognition (for reviews see Evans, 2007a; Evans & Stanovich, 2013) posit a qualitative difference between two types of processes: intuitive, resource-frugal processes (sometimes called "Type 1" or "System 1"); and analytic, effortful processes ("Type 2" or "System 2"), which draw heavily on attentional and working memory resources. Dual process approaches have become a mainstay of the New Paradigm in psychology of reasoning (Elqayam & Over, 2012; Oaksford & Chater, 2012, 2014). We will focus in particular on hypothetical thinking theory (Evans, 2006, 2007a), a dual process theory that suggests that hypothetical, effortful thinking is mainly invoked in novel situations which call for mental simulation of possibilities. Type 1 processes focus attention on the most relevant possibility or (epistemic) mental model (*relevance principle*), and only one model (*singularity principle*). The ensuing mental representation is accepted as default (*satisficing principle*), unless there is a good reason to reject it—in which case Type 2 processes get involved in revising or rejecting the default.[4]

According to the psychological suppositional conditional (Evans & Over, 2004), the Ramsey test involves both Type 1 and Type 2 processes. The most prominent Type 1 process involved in it is the *if-heuristic* (an idea going back to Evans, 1989). This is a special case of the relevance principle, in which the word "if" provides a relevance cue which focuses attention on the possibility that the antecedent is true (see also Skovgaard-Olsen et al., 2016). The if-heuristic thus provides the processing account of the attentional effects triggered by the Ramsey test. This gives rise to the defective truth table, because the attentional focus on true antecedent cases renders false antecedent cases irrelevant. Type 2 processes are invoked when the conditional triggers hypothetical thinking, that is, mental simulation of possibilities. The relevance principle will always play a role in generating the mental model, but Type 2 processing might be involved as well in evaluating it, especially when the situation is novel.

We propose that, for conditionals, *the relevant mental representation is by default the one in which there is an inferential relation between antecedent and consequent*. As argued in Krzyżanowska et al. (2014), this inferential link from antecedent to consequent need not be a deductive one. We propose a psychological mechanism: According to the satisficing principle, *the link need only be strong enough, in the sense of being subjectively supported*. For example, it can be supported by informal argumentation such as described by Hahn and Oaksford (2007); (see also Corner, Hahn, & Oaksford, 2011) by heuristic or pragmatic cues, as suggested by Evans and Over (2004); or by some form of inference to the best explanation (Douven, 2013, 2017a, 2017c; Douven & Mirabile, 2018; Douven & Schupbach, 2015). On the computational level, the semantic output of these cues often takes the shape of inductive or abductive inference, or even deductive inference. When relevance cues fail, the result is a truth value gap, accounting for the defective truth table. In this regard, our proposal echoes the psychological suppositional conditional; what our theory adds is the two novel hypotheses, that relevance takes the shape of inferential connection, and that the strength of this connection is bounded by satisficing. We will call these "the principle of relevant inference" and "the principle of bounded inference," respectively.

On the computational level, our proposal is compatible with a truth value gap semantics, in which a conditional is true if there is a strong enough argument from antecedent (plus background knowledge) to consequent; false if there is an argument connecting antecedent and consequent, but the argument is weak, or there is an argument (perhaps only a weak one) from the antecedent to the negation of the consequent; and neither true nor false if there is no inferential connection at all. (See Douven, 2016b, chap. 2, for more on this.) We note that, purely on the semantic level, inferentialism needs to accommodate the defective truth table anyway, in order to achieve descriptive adequacy. Once again, whether inferentialism, thus expanded, is tenable as a semantic theory we leave as a question for future research.

5. Experimental paradigm and hypotheses

Our theory construes the defective truth table as, in part, a product of failed inferential relevance between antecedent and consequent. It follows that, when a relevant inferential connection between antecedent and consequent is certain to exist, the patterns associated with the defective truth table should disappear or at least be substantially attenuated. To test our hypotheses, we therefore created a novel experimental task, the *soritical truth table task*, in which inferential relevance is guaranteed. Imagine that you are given a soritical series of color patches, the patches being numbered 1 through 14 and ordered from left to right. The series begins

[4] Although we only explicitly draw on the relevance and satisficing principles, these principles work in tandem with the singularity principle. Only a single relevant representation is generated, and people satisfice by sticking to this single model unless compelled to replace it.

Fig. 5.1. Soritical color series.

with a clearly blue patch—patch number 1—on the left, then gradually becomes more greenish as one progresses to the right, with adjacent patches being almost indistinguishable in color, and it ends with a clearly green patch—patch number 14—on the right (see Fig. 5.1).

You are now given a conditional sentence describing a relation between two of these patches, for example, "If patch number 6 is green, so is patch number 9." The positions in the series of the antecedent and consequent patches (6 and 9, respectively) jointly establish the "direction" of the inferential connection: in this case, because the soritical series goes from blue on the left to green on the right, the direction is (what we call) congruent, making the conditional true, according to inferentialism. It does not matter that patch number 6 is actually blue—that is, that the antecedent is false. The antecedent is still relevant because it provides necessary information about the direction and distance of the inferential connection. More generally, what we are asked to assume about any of the patches will, given what we know about the soritical series, be relevant to what we may infer about the color of the other patches (albeit with varying degrees of relevance). This is what we mean by "guaranteed inferential relevance"—that what is asserted in the antecedent is always relevant to the question of the truth of the conditional, regardless of whether the assertion is true.

The experiments we will report used soritical tasks like this one to elicit truth table judgments from participants. Participants were given the usual three response options, "True"/"False"/"Neither true nor false." We varied the distance (either adjacent or removed) and direction (either congruent or incongruent) of the consequent patch relative to the antecedent patch. In Experiment 1, for control purposes, we used three different presentation modes: verbal description only, visual presentation of soritical series throughout the test, and visual presentation of series through explanation but not during evaluation. Experiment 2 used a similar task, but with a different soritical series; Experiment 3 combines the two soritical series employed in Experiments 1 and 2; and in Experiment 4 we added a separate inference strength task. HIT generates several testable and novel predictions for this experimental paradigm, which we now specify.

5.1. If-heuristic override hypothesis

The soritical truth table task is designed to override the if-heuristic, because the context provides a powerful relevance cue to the contrary. Based on the principle of relevant inference, we hypothesize that the view or description of the series, aided by a soritical uncertainty, will focus participants' attention on the relationship between the antecedent and consequent patch, overriding the if-heuristic and hence directing participants' attention away from the truth value of the antecedent. If this is correct, we should see a minimal occurrence of defective truth table patterns in our data, and overall a very low prevalence of "Neither true nor false" responses. We explicitly test the if-heuristic override hypothesis in Experiment 3.

The if-heuristic is not without its critics; for example, Oaksford and Stenning (1992) argue that it is descriptive rather than explanatory. Here we use the term if-heuristic in the relatively uncontroversial sense of an attentional cue which focuses attention on the true-antecedent cases and away from the false-antecedent ones. We take this up again in the General Discussion.

5.2. Inferential strength hypothesis: The effects of distance and direction

HIT predicts that people judge a conditional to be true when they can satisfice on a strong enough argument leading from antecedent to consequent (relative to their background knowledge). To see what the requirement of an inferential connection amounts to specifically for the conditionals that we used as stimuli, note that there are two parameters that allow us to make an inference from the antecedent to the consequent, to wit, direction and distance. For instance, consider:

(6) a. If patch number 8 is green, so is patch number 11.
b. If patch number 8 is green, so is patch number 7.

Pre-theoretically, these conditionals strike us as true. And from an inferentialist perspective, they certainly are: (6a) because patches to the right of any given patch are greener than that patch, so the conditional is in the "right" (i.e., congruent) direction; and (6b) because patches number 8 and number 7 are adjacent, and so—given that adjacent patches differ almost imperceptibly in color—whatever is true for the color of patch number 8 can reasonably be expected to be true for the color of patch number 7.

There is a difference, however, between direction and distance.[5] Direction is a deductively valid cue: given what we know about the series, it is enough to know that patch number 8 is green to infer that patch number 11, which is in the congruent direction, is also green. By contrast, distance is a non-deductive cue, a parameter related to informal argument strength. In the case of (6b), distance is a strong cue, given that, within the series, adjacent patches have very similar colors. Specifically, on the supposition that patch number 8 is green, we have a strong warrant to believe that patch number 7 is green, too, even though the warrant is not as strong as in the case of (6a), in which the inferential connection is deductive in nature. Still, the non-deductive cue in (6b) is strong enough to satisfice on. Hence, argument strength in our experimental paradigm is operationalized as distance and direction effects—deductive

[5] We thank Rakefet Ackerman for drawing our attention to this distinction.

and non-deductive argument strength parameters, respectively.

Therefore, we expected direction to exert a main effect on truth evaluations of the conditionals: the consequent patch would either be positioned congruently, in the direction of the named color, strengthening the inference; or incongruently, away from the named color, weakening the inference. We also expected a main effect of distance, working as a probabilistic cue, such that the closer the distance between the patches, the stronger the inference, and hence the greater the prevalence of "True" responses.

Finally, we had an exploratory hypothesis for an interaction of direction with distance. If deductive and non-deductive cues interact, we could expect, for conditionals with the consequent patch on the congruent side, more "True" responses the greater the distance between antecedent and consequent, whereas the opposite should be observed for conditionals whose consequent patch is on the incongruent side. We left this hypothesis open as exploratory because we had no grounds to expect either an interaction between deductive and non-deductive cues, or its lack thereof.

5.3. Belief bias hypothesis

If the relation between antecedent and consequent is inferential, we would also expect it to be sensitive to the same psychological patterns that affect inferential processes in general. One of the most prominent and well-documented effects is belief bias, the tendency to be influenced by prior belief when drawing an inference, regardless of validity (Evans, Barston, & Pollard, 1983; Klauer, Musch, & Naumer, 2000). Prior belief has a robust effect on both deductive and non-deductive inference (Thompson & Evans, 2012); in many if not all cases, this effect centers on the believability of the conclusion: ceteris paribus, arguments whose conclusions are deemed believable are more often endorsed than arguments whose conclusions are deemed unbelievable.

Recall that HIT postulates an inferential connection, with the contextualized antecedent playing the role of a premise of an argument and the consequent playing the role of its conclusion. If participants in our task are susceptible to belief bias, we would expect them to base their inference at times on the truth value of the consequent, regardless of the inferential connection between antecedent and consequent—the analogue to belief bias in our experimental paradigm. For example, consider this conditional:

(7) If patch number 13 is green, so is patch number 11.

It is sufficient that patch number 11 is close to the green end of the series—no real inference is required, just considering the consequent patch position. Hence, we also expected an effect of consequent rank (i.e., the position of the consequent patch in the soritical color series) on the probability of judging a conditional to be true.

Belief bias is typically made up of three effects: a main effect of argument validity (or, in the case of informal reasoning, argument strength); a main effect of believability; and a belief × validity interaction, in which belief affects invalid (or weak) inferences more. If the analogy holds, we should expect an interaction between believability as measured by consequent rank, and the two argument strength parameters, distance and direction. Specifically, we would expect consequent rank to have a stronger effect where distance and direction provide no reliable cues to argument strength, as in the example above. (Note that this prediction is not as strong as the prediction for a main effect of consequent rank, as at least one study—Thompson and Evans, 2012—found no interaction effect for informal reasoning tasks.)

6. Experiment 1

6.1. Method

PARTICIPANTS

Seven hundred and four participants were recruited for a modest fee via the crowd-sourcing platform CrowdFlower (http://www.crowdflower.com), which directed them to the experiment on the Qualtrics platform (http://www.qualtrics.com). All participants were from Australia, Canada, the United Kingdom, or the United States. Data from participants who did not complete the study as well as from nonnative speakers of English were excluded from the analysis. This left us with 588 participants.

Participants in the visual presentation conditions were asked to classify the color of the patches (more details below). The participants who were asked to classify the color ($N = 397$) spent on average 381 s on the study (SD: 30 s); the participants in the description condition who did not have to classify the color ($N = 191$) spent on average 760 s on the study (SD: 408 s). (There is nothing surprising to the fact that the participants in the visual presentation conditions, which had to answer an *extra* question, spent on average *less* time on the study. The visual presentation of the soritical color series presented in Fig. 5.1 made the task easier.) For the analysis, we excluded from each group the fastest 5 percent responders as well as the slowest 5 percent responders. This left us with 532 participants: 359 in the visual presentation conditions, and 173 in the description condition. Of these 532 participants, 427 had a university education, while 105 had only a high school or secondary school education. The mean age of the participants was 34 years (±13). The remaining analysis and description is based on those 532 participants. The attrition rate (24 percent) is not unusual for web-based studies.

DESIGN

We used a 3 × 2 × 2 between-participants design with three levels of color series *presentation* ("description," "in sight," and "out of sight"), two levels of consequent *spread* ("small" and "large"), and two levels of *color* ("blue" and "green"), resulting in twelve groups in total. Participants were randomly assigned to the groups. The number of participants per group is given in Table 6.1. Each participant judged 22 conditionals, described in more detail below, with different values of direction and distance.

Table 6.1
Number of participants per between-participants condition.

		Presentation		
		Description	In sight	Out of sight
Spread	Small	49/37	45/55	47/44
	Large	43/44	39/41	46/42

Note. The first value in each cell denotes the number of participants in the blue color condition, the second value (i.e., after the slash) the number of participants in the green color condition. (See the text for an explanation of the factors.)

The *color* condition determined the value of X in schematic sentence (8) below (see Materials section) and was consistent for all conditionals. Thus, participants in the green condition were presented with conditionals which consistently referred to green, and participants in the blue condition were presented with conditionals which consistently referred to blue.

The 22 conditionals per participant resulted from a not fully-orthogonal combination of three within-subject factors *antecedent*, *direction*, and *range*. We used six different values for the antecedent i in (8): 2, 7, 8, 9, 10, and 13. For each of those antecedents, we presented either three (for antecedents 2 and 13) or four (for the remaining antecedents 7, 8, 9, and 10) different consequent values j. More specifically, for the latter four antecedents we presented two consequents in the congruent direction (the consequent patch to the left, bluer side of the antecedent patch in the blue condition and to the right, greener side of the antecedent patch in the green condition) and two consequents in the incongruent direction (the consequent patch to the right of the antecedent patch in the blue condition and to the left of the antecedent patch in the green condition). For one of the congruent patches in each direction the range was "near" whereas for the other the range was "far." The near range was always 1 step away. The value of the far range depended on the spread between-participants condition. In the small spread condition, the far patch was 2 steps away, whereas in the large spread condition, the far patch was 3 steps away. Note that for the two outer antecedents (i.e., 2 and 13) the far patches could not be realized for all directions (e.g., a patch to the left of 2 that is either 2 or 3 steps away would be outside of the color series).

To illustrate the design, three of the twelve groups—participants in the green color and small spread conditions—received conditionals of the forms "If patch number i is green, so is patch number $i - 1$," "If patch number i is green, so is patch number $i - 2$," "If patch number i is green, so is patch number $i + 1$," and "If patch number i is green, so is patch number $i + 2$," for $i \in \{7, 8, 9, 10\}$; for $i = 2$, they received conditionals of the first, third, and fourth forms (i.e., $i - 1$, $i + 1$, and $i + 2$); and for $i = 13$, they received conditionals of the first, second, and third form (i.e., $i - 1$, $i - 2$, and $i + 1$). The 22 conditionals were presented on the same screen in an individually randomized order.

The *presentation* condition determined how participants were presented with the color series. Close to one third of the participants—the participants in the description condition—only received the following description:

> Imagine a series of 14 color patches, numbered 1 through 14, and ordered from left to right. The series begins with a clearly blue patch—patch number 1—on the left. The patches then gradually become more greenish as we progress to the right, with adjacent patches being almost indistinguishable in color. The series ends with a clearly green patch—patch number 14—on the right.

Participants in the two visual presentation conditions—the in-sight and out-of-sight conditions—were shown the series of color patches displayed in Fig. 5.1. For the participants in the in-sight condition, the series was left in sight while they evaluated the conditionals. The participants in the out-of-sight condition were shown the series at the beginning but it was no longer in sight when they evaluated the conditionals; there was also no possibility for those participants to return to the screen with the color series. The series was presented in constant order, always from the blue left to the green right.

MATERIALS AND PROCEDURE

All materials were in English, the participants' native language, and shown on screen. Participants were asked to evaluate conditionals about the series of fourteen color patches shown in Fig. 5.1. The colors of the patches in that series were chosen along a constant line of lightness $L = 30$ and such that there is a subjective separation between adjacent patches of $\Delta E^* = 11.2$ as measured in CIELUV coordinates (see Fairchild, 2013, for details). The conditionals that participants were asked to judge were all of the form

(8) If patch number i is X, so is patch number j.

Participants were presented with three response options to judge the truth of each conditional, "True," "False," and "Neither true nor false."

All 359 participants who had been shown the series of color patches (i.e., participants in the in-sight and out-of-sight conditions) were asked to classify the colors in the soritical series in a separate task, after they had evaluated the conditionals. They were shown the color series again, and on the same screen were asked to indicate of each patch whether it was blue, green, or borderline blue/green, again in an individually randomized order. The responses to this question are displayed in Fig. 6.1.

STATISTICAL ANALYSIS

To test the predictions from HIT, we used two variables, *direction* and *distance*, where direction had two levels, "congruent" and "incongruent," and distance was determined by spread and range: When range was near, distance was always 1, independent of the value of the spread factor; when range was far, distance was either 2 (if spread was small) or 3 (if spread was large). Thus, distance

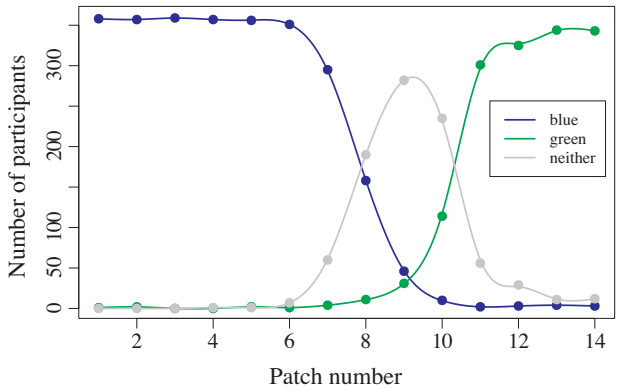

Fig. 6.1. For each patch, the number of participants that classified the patch as blue, green, and borderline. (Intermediate values have been interpolated and connected with a smooth line.)

varied both within and between participants, as range was varied within participants (i.e., participants saw both items in near and far range) and spread between participants (i.e., it was either small or large). As we did not specifically predict a linear or quadratic trend of distance, we treated distance as a categorical variable. For reasons given below, we compared HIT with models predicting effects of either consequent alone or of antecedent and consequent together. For these models, we used numeric variables with the patch number of antecedent and/or consequent centered at the midpoint of the scale.

Including all four independent variables describing the within-participants design (i.e., direction, distance, antecedent, and consequent) in one analysis was not possible as they are collinear; formally speaking, the matrix of a model with all four independent variables would be rank deficient. However, both a model with antecedent and consequent as independent variables and a model with direction, distance, and either antecedent or consequent as independent variables is perfectly possible. In addition to this, either type of model contains different information in the independent variables. Choosing between antecedent and consequent for the latter model only leads to a different parameterization of an otherwise equivalent model. Consequently, the results section is split into two parts: in the first, we compare the HIT model with two independently motivated models employing a model selection approach (e.g., Zucchini, 2000); in the second part, we investigate the HIT model further to assess whether the specific predictions are supported.

We analyzed our data using two binomial variables. In a first step, we only considered "Neither true nor false" versus other responses. In a second step, we only considered the other responses from the first step analyzing "True" versus "False" responses (i.e., excluding trials with "Neither true versus false" responses). In this way, we transformed a multinomial variable with three categories into two binomial variables (i.e., we analyzed the data as nested dichotomies; Fox, 2008). However, our design presented another statistical challenge. The models we compared all predict effects of within-participants variables such as antecedent or consequent patch or direction, which prohibits the use of standard statistical procedures for binomial variables such as logistic regression or χ^2-tests. These standard procedures assume independent and identically distributed responses, an assumption violated for within-subject factors. To overcome this problem, we employed an analysis based on generalized linear mixed models (GLMM; e.g., Jaeger, 2008), a type of repeated-measures logistic regression (see the supplemental materials for details).

We relegate a thorough discussion of the presentation factor (with levels "description," "in sight," and "out of sight") to the supplemental materials, as its effect on the results did not affect the conclusions. Our analysis was performed on the $532 \times 22 = 11{,}704$ individual responses.

6.2. Results and discussion

6.2.1. Indeterminate responses

Overall, participants classified 49.1 percent of the conditionals as true, 40.4 percent as false, and 10.5 percent as neither true nor false. However, the aggregate value for "Neither true nor false" does not adequately reflect the interindividual variability for the indeterminate responses, as 55 percent of the participants *never* chose this response option. Fig. 6.2 displays the distribution of the individual response proportions showing this clearly. Furthermore, only one out of 532 participants always responded with "Neither true nor false" (four participants always responded with "True" and five always with "False"). The results are even more striking when we examine specifically the false antecedent cases, the cases traditionally evaluated as indeterminate in classic truth table tasks. This is possible for the participants in the visual test conditions, who were also asked to classify the color patches as "green," "blue," or "borderline." Out of 1915 responses to conditionals with false antecedents, only 116 (6 percent) were indeterminate. These results strongly agree with our if-heuristic override hypothesis.

6.2.2. "True" versus "False" responses

Our main interest was the predictions concerning the rates of "True" versus "False" responses. Mean response proportions as a function of the independent variables relevant for HIT are displayed in Fig. 6.3. An eyeball test seems to confirm an effect of both inference strength and belief bias in line with the predictions of HIT.

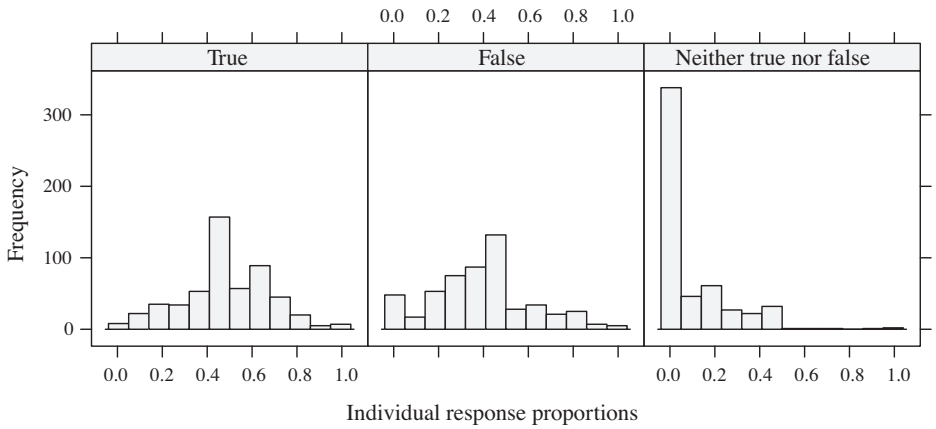

Fig. 6.2. Histogram of individual response proportions of Experiment 1.

Model selection. We started by fitting a GLMM for HIT, which included the variables deemed relevant: fixed effects for direction, distance, and consequent, plus all their interactions. We also fitted a consequent-only model, to represent pure belief bias effects, and an antecedent–consequent model, to represent a generic non-inferential semantic approach, based purely on the truth values of antecedent and consequent (see Section 2). For a fair comparison, all models contained fixed effects for the control variables color,

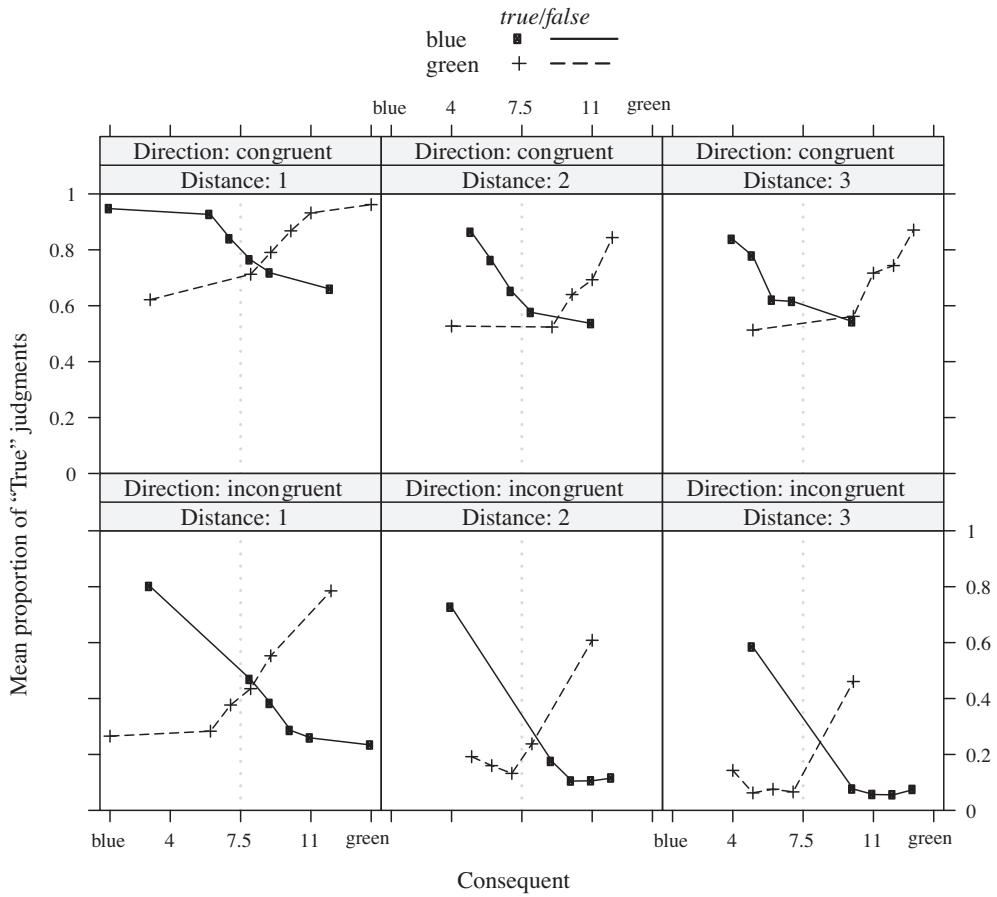

Fig. 6.3. Data relevant to HIT predictions concerning "True" versus "False" judgments (excluding all "Neither true nor false" responses) as a function of consequent (on the *x*-axis), distance between antecedent and consequent (increasing across columns from left to right), and direction (consequent patches on the congruent side are displayed in the upper panels, consequent patches on the incongruent side in the lower panels). Due to the mixed within-/between-participants nature of distance, the leftmost column with distance 1 is based on the same number of data points as the two other columns combined. (The supplemental materials contain a version of this figure that includes the "Neither true nor false" responses [Fig. 1] and versions that separate the presentation modes [Figs. 2–4].)

Table 6.2
Model comparison of GLMMs on "True" versus "False" responses for Experiment 1.

Model	K_f	K_r	LL	AIC	ΔAIC	BIC	ΔBIC
HIT	72	8/28	−3746.02	7708.03	0.00	8491.77	0.00
Consequent	24	2/1	−5526.32	11106.64	3398.61	11302.58	2810.80
Antecedent–consequent	48	4/6	−4156.75	8429.49	721.46	8850.39	358.62

K_f is the number of fixed effect parameters, K_r the number of random effect parameters (number of random slopes + random intercept/number of correlations among random effects), and LL the maximum log-likelihood of each model. AIC is the Akaike Information Criterion and BIC the Bayesian Information Criterion, two indices for model selection that take model fit (i.e., $-2 \times$ LL) and model complexity (i.e., number of estimated parameters, and for BIC also sample size) into account. ΔAIC and ΔBIC are the values for each model minus the smallest AIC or BIC value. Models with smaller indices provide a more parsimonious (i.e., better) description of the data.

presentation, and spread (where possible) plus full interactions with the relevant variables (note that the presence or absence of these variables did not affect the conclusions).

Model selection results are displayed in Table 6.2 and were straightforward. As could be expected from the descriptive results, the HIT model clearly provided the best account in terms of both AIC and BIC.[6] While the antecedent–consequent model provided the second best account, its performance was dramatically worse, with ΔAIC = 721.

Analysis of HIT model. Next we tested the specific predictions of HIT. To this end, we estimated *p*-values for all effects of the HIT model (the full results can be found in Table 1 in the supplemental materials). The inferential strength hypothesis was supported by a main effect of direction, $\chi^2(1) = 201.66, p < .0001$, indicating that conditionals with consequent patches on the congruent side were almost unanimously judged to be true (estimated marginal mean on the response scale [EMM] = .93),[7] whereas conditionals with consequent patches on the incongruent side were judged to be true in less than 20 percent of the cases (EMM = .18). We found the predicted main effect of distance, $\chi^2(2) = 80.00, p < .0001$, indicating that conditionals with consequent patches one step away (EMM = .86) were more likely to be judged "True" than conditionals with consequent patches two steps away (EMM = .60, odds ratio [OR] = 4.26, $z = 6.24, p < .0001$), which in turn were more likely to be judged "True" than conditionals with consequent patches three steps away (EMM = .35, OR = 2.70, $z = 4.46, p < .0001$). The exploratory prediction of interaction of direction with distance was not supported, $\chi^2(2) = 0.50, p = .78$.

We also found strong support for the predicted belief bias effect. First, we found the analogue of the main effect of belief, an effect of consequent (i.e., an interaction of consequent with color), $\chi^2(1) = 356.10, p < .0001$. The slope in the blue condition was clearly negative, $b_{blue} = -0.86$, 95% CI [−0.99, −0.73], whereas the slope in the green condition was clearly positive,$b_{green} = 0.71$, 95% CI [0.59, 0.83]. Second, we also found evidence for an analogue of the believability × validity interaction, namely, that the effect of consequent was stronger for weaker inferences as indicated by a significant three-way interaction of direction × color × consequent, $\chi^2(1) = 7.69, p = .006$. Follow-up analyses further confirmed the predictions: the slopes for consequent in the congruent conditions ($b_{blue} = -0.76$, 95% CI [−0.92, −0.59], and $b_{green} = 0.61$, 95% CI [0.47, 0.75]) tended to be smaller than the slopes in the incongruent conditions ($b_{blue} = -0.97$, 95% CI [−1.13, −0.81], and $b_{green} = 0.80$, 95% CI [0.65, 0.95]), both $|z| > 2.1$, both $p = .055$. Fig. 6.4 shows the fixed and random effects model estimates of this interaction which reveal that, although there is considerable individual variation, the pattern was quite consistent.[8] Given the absence of the direction × distance interaction, the absence of the direction × distance × color × consequent interaction, $\chi^2(2) = 3.09, p = .21$, was entirely unsurprising.

7. Experiment 2

Consider the following sentences:

(9) a. If Dino wins the lottery, he will quit his job.
b. If Dino were to win the lottery, he would quit his job.
c. If Dino had won the lottery, he would have quit his job.

Of these sentences, (9a) is ordinarily classified as an indicative conditional while (9b) and (9c) are classified as subjunctives, with (9c) counting as a special kind of subjunctive, usually called "counterfactual," given that it pragmatically implicates its antecedent (and in

[6] Because we do not assume that the true model is among our candidate models, we have a preference for AIC over BIC (see Yang, 2005).

[7] Marginal means were estimated at the midpoint of consequent (i.e., at 7.5), *p*-values of follow-up tests were corrected for each significant effect separately using a generalized version of the Bonferroni–Holm method (see supplemental materials for details).

[8] Taking a step back from the predictions, inspection of Fig. 6.4 also reveals a main effect of color, $\chi^2(1) = 15.47, p < .0001$, as the *y*-axis position with which the midpoint of consequent is crossed differs between the blue and the green condition. In other words, the probability of a "True" response at the midpoint is higher in the blue (EMM = .74) than in the green condition (EMM = .50). This effect is a consequence of the perceived asymmetry of the color series in the visual presentation conditions, as is evident from Fig. 6.1. Hence, it can be explained by a color × presentation interaction, $\chi^2(2) = 13.29, p = .001$. Follow-up contrasts on the interaction revealed the to-be-expected pattern: While the difference between the blue and the green condition is significant in the two visual presentation conditions ($EMM_{in\text{-}sight}$ = .76 and .34; $EMM_{out\text{-}of\text{-}sight}$ = .73 and .35), both OR >5.1, both $p < .0006$, no such effect was observed in the description condition, in which no asymmetry could be expected (EMM = .74 and .79), OR = 0.74, $p = .51$. Note that this interaction also subsumed a main effect of presentation, $\chi^2(2) = 11.88, p = .003$.

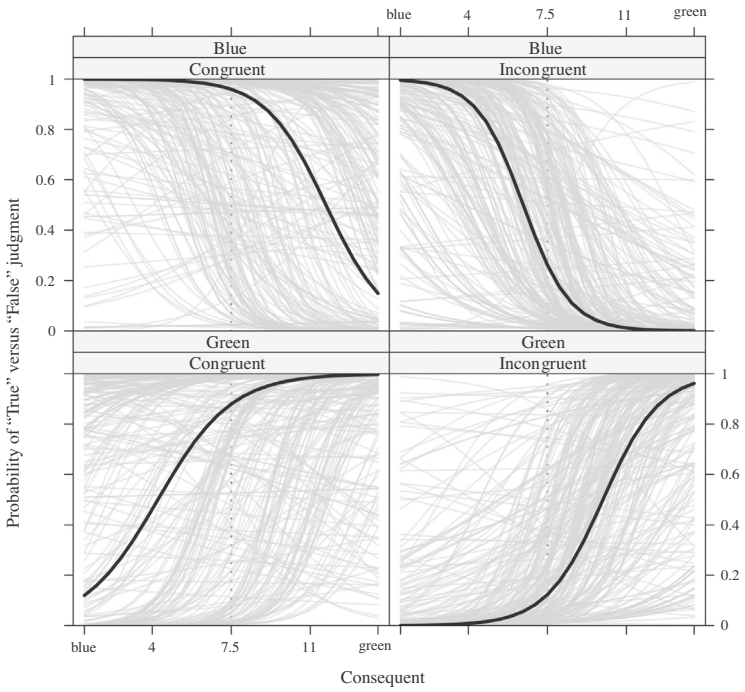

Fig. 6.4. Three-way interaction of direction × color × consequent of HIT model for "True" versus "False" responses. The black lines show the predictions based on the fixed effects. The gray lines in the background show the predictions of the individual random effects from all 532 participants; these are plotted with 50 percent transparency so that darker lines represent more participants.

fact also its consequent) to be false. As stated at the outset (note 1), the present paper is exclusively concerned with indicative conditionals.[9] However, consider

(10) If patch number 2 is green, so is patch number 3.

As can be seen in Fig. 6.1, virtually all participants in the in-sight and out-of-sight conditions judged patch number 2 to be *blue*; and the participants in the description condition, based on the information they were provided, will have had the reasonable expectation that this patch was blue. So, although (10), like all conditionals in our materials, is grammatically an indicative conditional, there is a legitimate concern that at least some participants will have read it and other conditionals with clearly false antecedents as counterfactuals and thus as subjunctives, and that this reading may have affected the outcomes of Experiment 1.

Experiment 2 was designed to address this concern, by using a soritical series which does not allow with any certainty the attribution of truth values to the antecedents or consequents of the conditionals concerning the series. Instead of a series of color patches, we used the series of spheres shown in Fig. 7.1, where the context provided only information sufficient to judge the *relative* rather than *absolute* sizes of the spheres; all conditionals we used were to the effect that if a given sphere in the series was large, then so was another given sphere in the series. The context did not support a counterfactual interpretation of these conditionals, given that neither the conditionals' antecedents nor their consequents could be said to be false with any degree of confidence. For example, although sphere number 14 is clearly larger than sphere number 13, it is impossible to say that it is *large*—for all we know, all 14 spheres are exceedingly small, having been produced by a miniaturist artist under a microscope. Nevertheless, the series still supports inferential connections between the antecedent and consequent of the relevant conditionals.

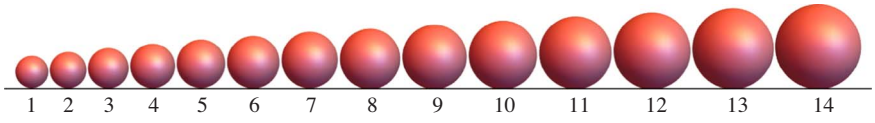

Fig. 7.1. Soritical series of spheres.

[9] As noted in Douven (2016b), it may not take too much effort to tweak inferentialism—one of the main pillars of HIT—to make it apply to indicative and subjunctive conditionals alike. Thus generalized, inferentialism might well serve as a foundation for a version of HIT that covers subjunctive conditionals as well. Here, however, we flag a possible extension of HIT in this direction only as an avenue for future research.

7.1. Predictions

Our predictions were similar to the ones we had in Experiment 1, albeit with some differences. We predicted a replication of the inferential strength effect, articulated as main effects of direction and distance. We left the distance × direction interaction as an exploratory hypothesis again. We also predicted a consequent effect (our belief bias analogue), although we expected it not to be as strong as in Experiment 1. This is because sphere size in this series is purely relative, and none of the spheres can be said to be small or large; hence, there is little belief *to* bias the inference. As a minor prediction, we predicted a consequent × direction effect, a replication of the same effect from Experiment 1. As before, this prediction is less firm because believability × validity interaction is not universal in belief bias.

7.2. Method

PARTICIPANTS

Fifty-six participants were recruited in the same manner as in Experiment 1. We excluded from analysis data from the 5 percent slowest and 5 percent fastest participants, then from non-native speakers of English, participants who did not have normal or corrected to normal vision, color blind or dyslexic participants, and participants who failed either of two validation questions. The first validation question (following Pennycook, Trippas, Handley, & Thompson, 2014) appeared at the end of the demographic section. Participants were given a list of hobbies and were asked, "Below is a list of hobbies. If you are reading these instructions please write 'I read the instructions' in the 'other' box." Data from participants who left the box empty or specified hobbies were excluded. The second validation question, taken from Aust, Diedenhofen, Ullrich, and Musch (2013) was placed at the end of the study, and asked participants to state if they had responded seriously to the questions in the experiment. We excluded data from participants who responded with a "No." Lastly, we excluded data from participants who participated in the study using a handheld device such as a smartphone (information about the device was obtained from their browser information, provided by Qualtrics). This left us with 39 participants. These participants spent on average 302 s on the experiment (SD: 100 s). Thirty of them had a university education and 9 had only a high school or secondary school education. Their mean age was 41 years ($\pm$11).

DESIGN AND MATERIALS

We used the soritical series of 14 spheres shown in Fig. 7.1. Participants were given the following instructions:

> At the top of the screen you see a series of 14 spheres. These spheres are all aligned, one lying next to the other. Imagine that you see the spheres from an unknown distance. They can be very far away or quite nearby, although all are the same distance from you. You do not know anything about the *absolute* size of these spheres.

We used a slightly simplified version of the design in Experiment 1, with only the in-sight visual condition, and (obviously) no manipulation of color. Antecedent spheres were in position 2, 7, 8, 9, 10 or 13; for each antecedent sphere, the consequents were $+/-1$ and $+/-3$ (with the exception of antecedent spheres 3 and 12, for which -3 and $+3$ were impossible, respectively). Participants were presented with the full set of 22 items on the same page in an individually randomized order.

7.3. Results and discussion

7.3.1. Indeterminate responses

Overall, participants classified 58.2 percent of the conditionals as true, 27.4 percent as false, and 14.5 percent as neither true nor false. Again, only one of the 39 participant always responded with "Neither true nor false" while the majority never used this response (two participants always responded with "True" and zero always with "False"). Fig. 7.2 displays the distribution of the individual response proportions. The pattern shown broadly replicates that of Experiment 1, supporting our if-heuristic override hypothesis. The rate of indeterminate responses was slightly higher in this experiment, perhaps because a sizable minority (12 participants) never used "False" as response.

7.3.2. "True" versus "False" responses

Our main analysis again concerned the rate of "True" versus "False" responses, as HIT predicted a unique pattern. Mean response proportions as a function of the independent variables are displayed in Fig. 7.3.

Eyeball inspection of the Figure reveals that the results conceptually replicate those of Experiment 1 (Fig. 6.3), although the pattern differs in some respects. On the one hand, consequent effects seem considerably weaker here; on the other hand, direction effects seem at least equally strong or even stronger. Furthermore, there seems to be evidence for a distance × direction interaction: the distance effect seems strong in the incongruent direction, but absent in the congruent direction.

Model selection. In the first step we compared again three GLMMs: a model for HIT (with fixed effects for direction, distance, consequent, and all interactions), a consequent model (with consequent as its only fixed effect), and an antecedent–consequent model (with antecedent, consequent, and their interaction as fixed effects). Results are displayed in Table 7.1. In line with the eyeball inspection of Fig. 7.3, the HIT model clearly provides the best account, $\Delta AIC > 22$.

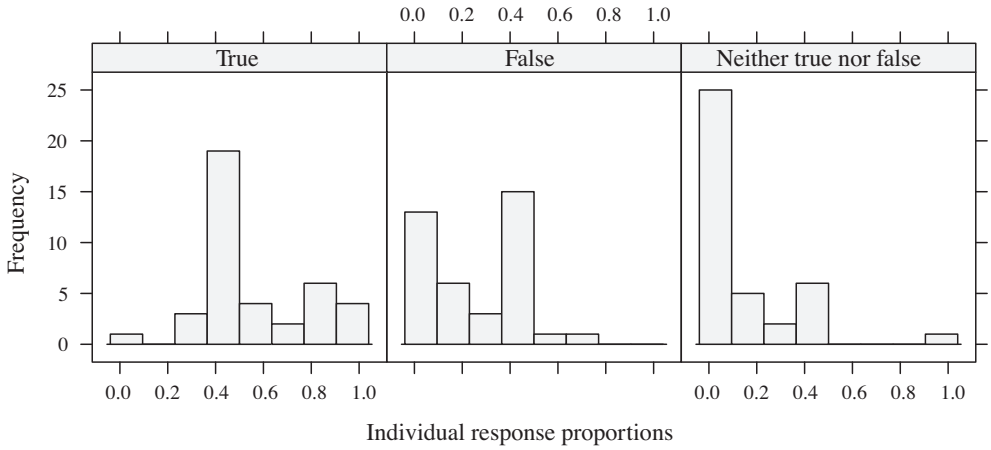

Fig. 7.2. Histogram of individual response proportions of Experiment 2.

Analysis of HIT model. Next we tested for the specific predictions of HIT.[10] In support of the inferential strength hypothesis, we found a very strong main effect of direction, $\chi^2(1) = 34.76, p < .0001$, OR > 5000, indicating that conditionals with consequent spheres on the congruent side were unanimously judged to be true (EMM = 1.00), while those on the incongruent side were only judged as true in about a fifth of the cases (EMM = .22). The effect of distance was also present, but less pronounced, $\chi^2(1) = 6.15, p = .01$, OR = 7.35. Consequent spheres one step away (EMM = .98) were more likely to be judged "True" than con-

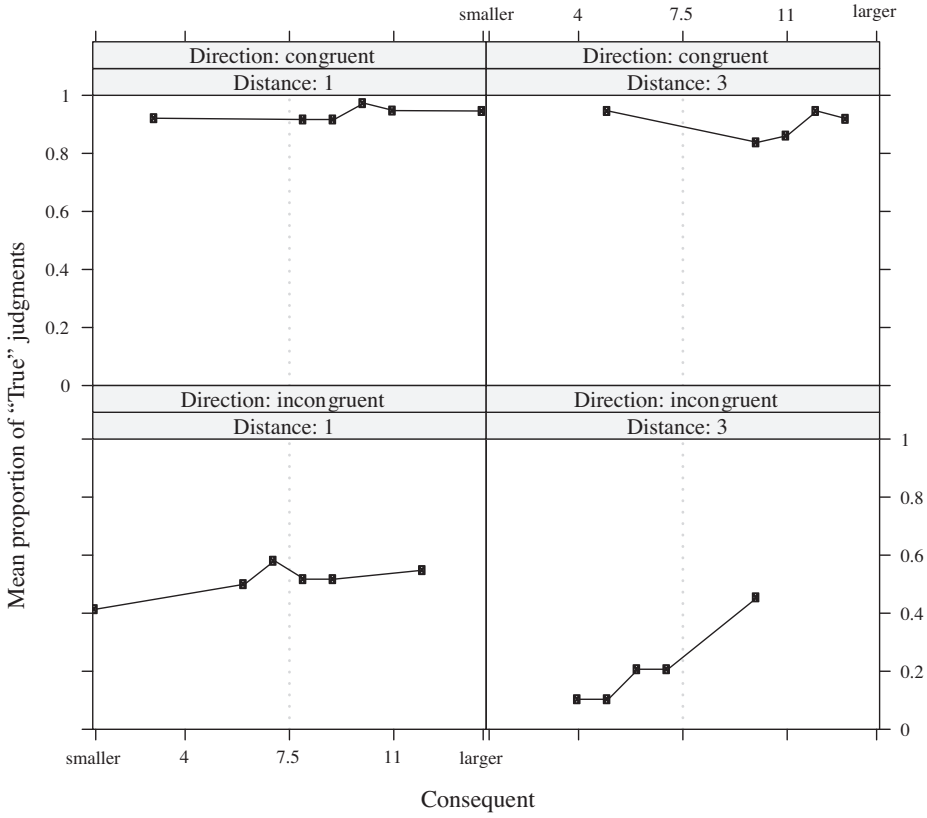

Fig. 7.3. Data from Experiment 2 relevant to HIT predictions concerning "True" versus "False" judgments (excluding all "Neither true nor false" responses) as a function of consequent (on the *x*-axis), distance between antecedent and consequent, and direction. (Fig. 8 in the supplemental materials is a version of this figure that includes the "Neither true nor false" responses.)

[10] This analysis is based on the HIT model without correlation among random slopes as we were unable to reliably obtain *p*-values for the model including correlations (see, e.g., Bates, Kliegl, Vasishth, & Baayen, 2015). See supplemental materials Table 3 for full results.

Table 7.1
Model comparison of GLMMs on "True" versus "False" responses for Experiment 2.

Model	K_f	K_r	LL	AIC	ΔAIC	BIC	ΔBIC
HIT	8	8/28	−178.46	444.93	0.00	647.26	115.07
Consequent	2	2/1	−380.99	771.99	327.06	794.98	262.79
Antecedent–consequent	4	4/6	−219.91	467.82	22.89	532.20	0.00

Note. See Table 6.2. The apparent better performance of the antecedent–consequent model compared to HIT in terms of BIC is a consequence of the correlation parameters among random effects and the large penalty provided by BIC for each parameter. After removing those correlations, HIT provides the best account in terms of both AIC, ΔAIC > 150, and BIC, ΔBIC > 110. This suggests that, given the modest sample size, estimating correlations among random slopes is not completely justifiable (see also Bates et al., 2015).

sequent spheres three steps away (EMM = .89). As in Experiment 1, we did not find strong evidence for the direction × distance interaction, $\chi^2(1) = 3.32$, $p = .07$ (but see below).

We again found support for the predicted belief bias effect, but as expected, it was not as strong as the effect found in Experiment 1, $\chi^2(1) = 8.27$, $p = .004$, $b = 0.26$, 95% CI [0.10, 0.43]. Furthermore, we again found the analogue to the validity × believability interaction, the direction × consequent interaction, $\chi^2(1) = 10.61$, $p = .001$. The effect of consequent (i.e., "believability") was absent for congruent (i.e., "valid") spheres, $b = 0.03$, 95% CI [−0.18, 0.24], but clearly present for incongruent (i.e., "invalid") spheres, $b = 0.49$, 95% CI [0.25, 0.73].

In addition, we also found a three-way interaction of direction × distance × consequent, $\chi^2(1) = 10.90$, $p = .001$. This interaction is displayed in Fig. 7.4 and more clearly exhibits the pattern discussed above. The belief bias effect of consequent seems to appear only for "invalid and improbable" spheres, that is, those which are neither in the congruent direction nor only one step away. Specifically, the effect of consequent is only significantly above zero for incongruent spheres with distance 3, $b = 0.84$, 95% CI [0.43, 1.25]. For congruent spheres with distance 3, the effect of consequent is virtually zero, $b = -0.08$, 95% CI [−0.38, 0.22]. For spheres with distance 1, the estimated effect of consequent is $b = 0.14$ for both congruency conditions, with 95% CI [−0.14, 0.42] for

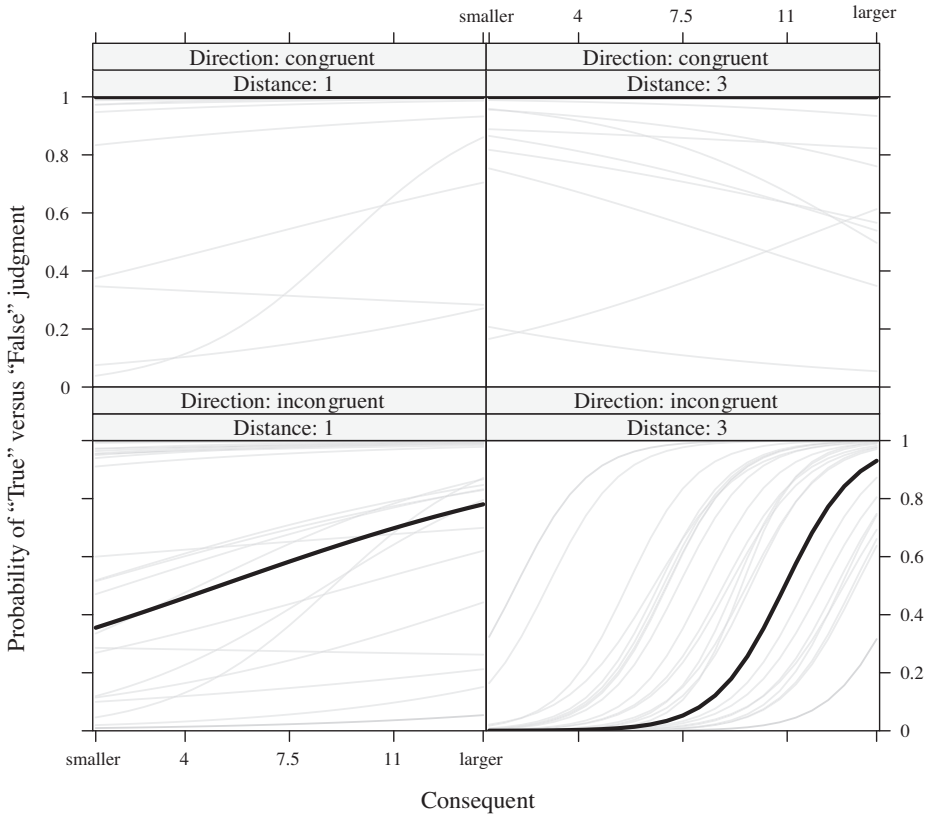

Fig. 7.4. Three-way interaction of direction × distance × consequent of HIT model for "True" versus "False" responses of Experiment 2. The black lines show the predictions based on the fixed effects. The gray lines in the background show the predictions of the individual random effects from all 39 participants; these are plotted with 50 percent transparency so that darker lines represent more participants.

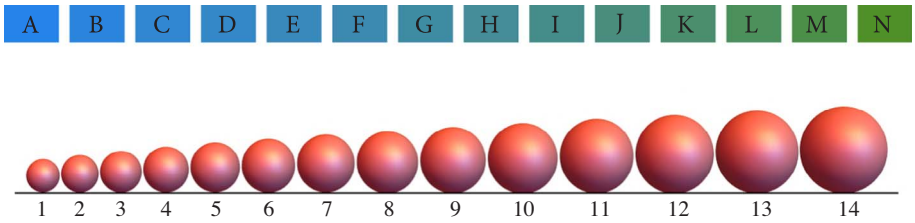

Fig. 8.1. Task structure for Experiment 3. Participants were shown the above while being given conditionals of two main types: conditionals with both antecedent and consequent within a single series (e.g., "If patch A is blue, then patch D is blue"; "If sphere 1 is large, then sphere 4 is large"); and conditionals in which the antecedent and the consequent were between the series (e.g., "If patch A is blue, then sphere 4 is large"; "If sphere 1 is large, then patch D is blue"). Relevance is guaranteed for within-series but not for between-series conditionals.

congruent spheres and 95% CI [−0.06, 0.34] for incongruent spheres. The reason for the apparent difference between the two distance 1 conditions is a different intercept (i.e., EMM). The EMMs for the four conditions are $EMM_{cong1} = 1.00$, $EMM_{cong3} = 1.00$, $EMM_{incong1} = .58$, and $EMM_{incong3} = 0.05$.

8. Experiment 3

According to our if-heuristic override hypothesis, the soritical truth-table task provides a strong cue to the guaranteed relevance of the antecedent even when the antecedent is false. This leads to a prediction of an unusually low proportion of indeterminate responses, compared to field benchmarks. Experiments 1 and 2 provided strong evidence for this hypothesis, with a very low proportion of indeterminate responses. However, Experiments 1 and 2 did not provide a control condition with no guaranteed relevance. The aim of Experiment 3 was to test our predictions against a control condition with a comparable task which nevertheless does not guarantee the relevance of the antecedent, providing a more robust test of the if-heuristic override hypothesis. Thus, Experiment 3 is a direct test of HIT's *principle of relevant inference*.

To create a control condition, we presented participants with the same soritical truth-table task, but instead of a single soritical series we combined the colored patches series used in Experiment 1 with the spheres series used in Experiment 2. (See Fig. 8.1.)

In the experimental conditions, participants were presented with conditionals whose antecedent and consequent were within a single series; for the control conditions, participants were presented with conditionals whose antecedent and consequent were between series. Thus, we had two main test conditions articulated through four types of conditionals:

1. Within-series conditionals (experimental condition; guaranteed relevance):
   (a) within-series colors–colors (e.g., "If patch A is blue, then patch D is blue");
   (b) within-series spheres–spheres (e.g., "If sphere 1 is large, then sphere 4 is large").
2. Between-series conditionals (control condition; no guaranteed relevance):
   (a) between-series colors–spheres (e.g., "If patch A is blue, then sphere 4 is large");
   (b) between-series spheres–colors (e.g., "If sphere 1 is large, then patch D is blue").

In the between-series control condition, in which participants had to infer from an antecedent in one series to a consequent in another series, relevance of the antecedent to the consequent was blocked. By contrast, the within-series experimental conditions required participants to infer within the same soritical series, so that relevance *was* guaranteed. Moreover, to suppress relevance more effectively, the direction in each series was contralateral: for the colors series the congruent direction was right-to-left (i.e., color terms always referred to "blue"), whereas for the spheres series the congruent direction was left-to-right (i.e., sphere terms always referred to "large"). Hence, participants could not draw inference from one series to another merely by analogy to the number of required steps in the other series. Thus, for the within-series condition we expected a replication of the pattern we identified in Experiments 1 and 2, whereas for the between-series condition we expected participants to revert to the defective truth-table pattern typically observed in the field.

8.1. Predictions

Our main prediction concerned the if-heuristic override. The manipulation we introduced was designed so that the between-series condition would suppress the strong relevance cue delivered by the use of the soritical series, while this cue would be largely preserved in the within-series condition. Furthermore, in the between-series condition we blocked participants from any quick-and-easy ways to infer from antecedent to consequent. Therefore, we expected participants to resort in that condition to the usual pattern observed for abstract conditionals, namely the defective truth table (TF##), perhaps with a minority conforming to the conjunctive pattern (TFFF) as found in previous studies (e.g., Evans et al., 2003; Oberauer & Wilhelm, 2003). By contrast, we expected the proportion of indeterminate responses in the within-series condition to remain comparable to that of the previous two experiments. The main prediction for Experiment 3 was therefore that the proportion of indeterminate responses would be significantly higher in the between-series condition relative to the within-series condition.

The defective truth table means that the rate of indeterminate responses in the between-series condition should be a function of the truth of the antecedent: where the antecedent is false or indeterminate, the conditional should be evaluated as indeterminate (e.g., Evans et al., 2003; Baratgin et al., 2013). As an illustration, see Fig. 6.1. As the antecedent progresses from left to right, patches are predominantly evaluated first as blue (= "True"), then as neither blue nor green (= "Indeterminate"), and finally as green (= "False"). Thus, we predicted that the proportion of indeterminate responses in the between-series conditions would increase as a function of antecedent rank when moving toward the "False" end of the scale—that is, toward the green end of the scale in the colors–spheres condition, and toward the smaller end of the scale in the spheres–colors condition.

We also had auxiliary predictions for the pattern of determinate responses. Broadly, we expected a replication of the pattern in Experiments 1 and 2 for the within-series condition: main effects of distance, direction, and consequent. However, even for the within-series condition, the situation is somewhat more complex relative to Experiments 1 and 2, because the second series is always in sight and might provide misleading cues. For the between-series condition, participants were likely to create ad hoc heuristics to reduce the cognitive load, but we had no way to know in advance what these heuristics might be. We therefore left predictions for the determinate responses in this experiment exploratory.

8.2. Method

PARTICIPANTS

One hundred and seventy-three participants were recruited the same way as in the previous experiments. We used the same validation measures and exclusion criteria as in Experiment 2, which left us with 116 participants. These participants spent on average 481 s on the experiment (SD: 158 s). Ninety of them had a university education and 26 had only a high school or secondary school education. Their mean age was 39 years (±11).

DESIGN AND MATERIALS

The experiment was a combination of Experiments 1 ("in-sight" condition) and 2. In each trial, participants always saw both the soritical color patches series on top and the soritical series of spheres right underneath it, as shown in Fig. 8.1. The soritical series of color patches was labeled from A to N instead of from 1 to 14 (as it was in Experiment 1), so that all patches and spheres would have different labels. Color terms always referred to "blue" and sphere terms always referred to "large".

Each participant was presented with a total of 24 conditionals in two main *relevance* conditions: 12 conditionals in the *within-series* condition (guaranteed relevance), in which the conditional referred to only one of the two soritical series; and 12 conditionals in the *between-series* condition (no guaranteed relevance control condition), in which the conditional referred to one of the soritical series in the antecedent and to the other one in the consequent. For the 12 conditionals in each condition the consequent patches/spheres were in position 4, 8, and 11 (or D, H, and K); for each consequent patch sphere, the antecedents were $+/-1$ and $+/-3$. This resulted in a balanced design for the HIT relevant factors distance and direction.

Between-participants we manipulated the type of soritical series in the within-series (1a or 1b) and between-series (2a or 2b) conditions. Participants were randomly assigned to one of four groups:

- ◇ Within-series conditionals: colors (condition 1a); between-series conditionals: colors–spheres (condition 2a); $N = 24$.
- ◇ Within-series conditionals: colors (condition 1a); between-series conditionals: spheres–colors (condition 2b); $N = 37$.
- ◇ Within-series conditionals: spheres (condition 1b); between-series conditionals: colors–spheres (condition 2a); $N = 25$.
- ◇ Within-series conditionals: spheres (condition 1b); between-series conditionals: spheres–colors (condition 2b); $N = 30$.

The levels of the direction variable for the within-series conditions were defined as in Experiments 1 and 2, respectively. Specifically, in condition 1a (colors–colors) consequent patches on the left (i.e., bluer) side of the antecedent were considered congruent (and incongruent otherwise) and in condition 1b (spheres–spheres) consequent patches on the right (i.e., larger) side of the antecedent were considered congruent (and incongruent otherwise). Thus, for example, the item "If patch E is blue, so is patch D" is congruent; similarly, the item "If sphere 5 is large, so is sphere 6" is also congruent.

Note that the congruent direction for the color patches series is from right to left, whereas the congruent direction for the spheres series is from left to right. As previously mentioned, this was a deliberate choice whose aim was to suppress relevance by severing connections between the series. Since congruency is defined by the relation between antecedent and consequent, and since this connection is deliberately disrupted in the between-series condition, this means that defining direction in the between-series conditions is less straightforward than in the within-series conditions. It is only possible to define direction analogously, either based on the antecedent patch or based on the consequent patch. For example, in the conditional "If patch E is blue, then sphere 6 is large", the congruent direction based on the consequent is left-to-right, whereas based on the antecedent it would be right-to-left. While either direction is equally philosophically plausible, psychological plausibility in this case calls for defining between-series conditionals based on the *consequent*. Recall that Experiments 1 and 2 found strong evidence for a belief bias analogue, that is, for a main effect of the consequent rank. Thus, in the soritical paradigm the consequent provides a strong heuristic cue to the truth value of the conditional. For the between-series conditions, then, we defined congruency analogously, based on the position of the consequent patch, so that congruency in each between-series condition was defined according to the analogous within-series condition with the same consequent. Thus, congruency in condition 2a (colors–spheres) was defined as it was in condition 1b (spheres–spheres), and congruency in condition 2b (spheres–colors) was defined as it was in condition 1a (colors–colors). For example, for the conditional "If patch E is blue, then sphere 6 is large", the congruent direction is left-to-right.

The order of the conditionals was individually randomized, with one conditional presented per page.

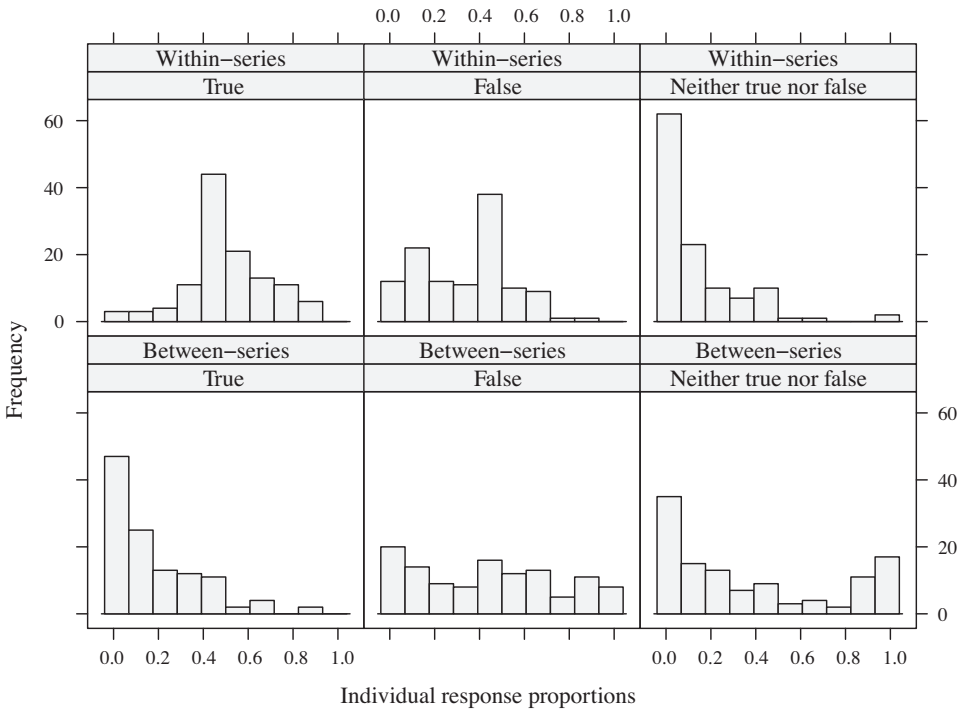

Fig. 8.2. Histogram of individual response proportions of Experiment 3. The upper row shows the response proportions for the within-series conditionals (which refer to the same soritical series in both antecedent and consequent, and hence guarantee relevance) and the lower row shows the between-series conditionals (which refer to different soritical series in antecedent and consequent, with no guarantee of relevance).

8.3. Results and discussion

8.3.1. Indeterminate responses

To test our if-heuristic override hypothesis, we first compared the rate of determinate versus indeterminate responses in the within-series versus between-series test condition. Fig. 8.2 displays the full distribution of the individual response proportions separated for within-series condition (upper row) and between-series condition (lower row). As expected, the rate of indeterminate responses in the between-series condition outnumbered the rate of indeterminate responses in the within-series condition almost 3:1. The results for the within-series condition replicated those of Experiments 1 and 2. The observed rate of indeterminate responses was 13 percent (52 percent "True" and 35 percent "False") and on the individual level 53 percent of participants never chose the indeterminate response (3 percent never chose "True" and 10 percent never chose "False"). For the between-series condition, the pattern was markedly different and, as expected, the proportion of indeterminate response replicated the typical defective truth table pattern (e.g., Schroyens, 2010): 38 percent "Neither true nor false" responses, 44 percent "False" responses, and 18 percent "True" responses. On the individual level, only 30 percent never chose the "Neither true nor false" response whereas 41 percent never chose "True" and 17 percent never chose "False". In addition, for the between-series condition 15 percent of participants always responded with the indeterminate response (7 percent always with "False" and 0 percent always with "True"), whereas this rate was only 2 percent for the within-series condition (0 percent always with "False" and 0 percent always with "True").

To investigate this pattern further, we estimated a GLMM with the rate of indeterminate responses versus other responses as dependent variable.[11] The independent variables (i.e., fixed effects) were *antecedent*, a numerical variable from 1 to 14, centered at the midpoint for the analysis; *type*, a mixed within–between factor with four levels derived from the group factor and the relevance factor: (a) within-series colors, (b) within-series spheres, (c) between-series colors–spheres, and (d) between-series spheres–colors; *direction*; *distance*; and their interactions. Recall that the defective truth table pattern means that the rate of indeterminate responses in the between-series condition should be a function of the truth of the antecedent. Thus, we predicted that the slope in the between-series condition should significantly differ from 0, and that the slope in the between-series conditions should be significantly steeper than the slope in the within-series conditions.

The GLMM revealed the expected main effect of type, $\chi^2(3) = 54.09$, $p < .0001$. Inspection of all pairwise comparisons of the four means confirmed the predictions: the rate of indeterminate responses was larger for the between-series conditionals (EMM = .24) than for the within-series conditionals (EMM = .01). Neither levels (a) and (b) nor levels (c) and (d) differed from each other, both

[11] All tests for fixed effects for Experiment 3 are based on a model without correlations among random parameters, due to numerical problems in obtaining *p*-values in the model with correlations.

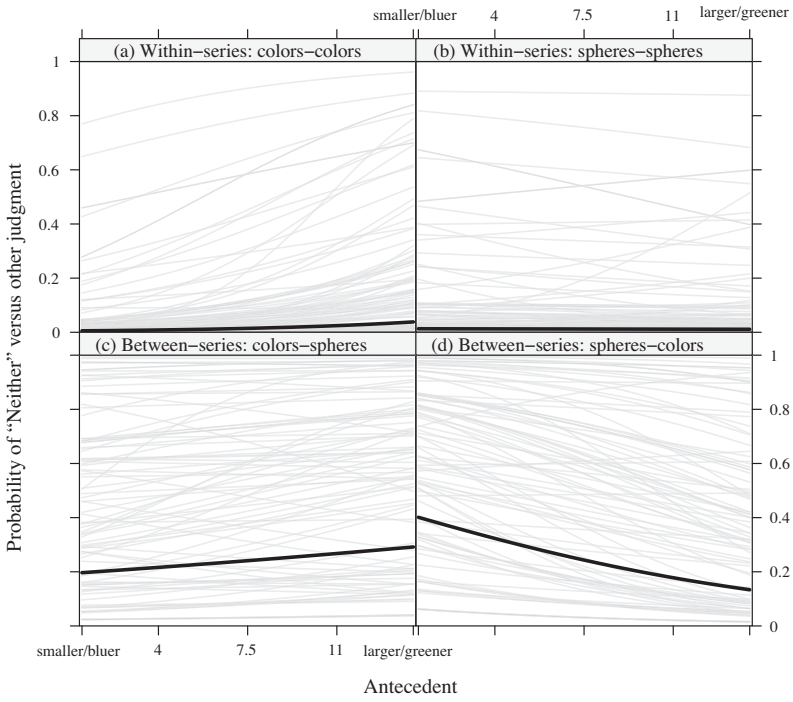

Fig. 8.3. Two-way interaction of type × direction for "Neither" versus other responses of Experiment 3. The black lines show the predictions based on the fixed effects. The gray lines in the background show the predictions of the individual random effects from all 116 participants; these are plotted with 50 percent transparency so that darker lines represent more participants.

ps > .98, whereas all other pairwise comparisons were significant, all ps < .0001. We also found a main effect of direction, $\chi^2(1) = 13.38$, $p = .0003$, which indicated that, overall, participants gave more indeterminate responses to incongruent items (EMM = .09) than to congruent items (EMM = .04). This main effect was qualified by a type × direction interaction, $\chi^2(3) = 32.16$, $p < .0001$, and a type × direction × distance interaction, $\chi^2(3) = 8.80$, $p = .03$. Follow-up analysis on the latter interaction revealed that the effect of direction was only significant for far patches in condition (a) ($EMM_{cong} = .01$ versus $EMM_{incong} = .05$, $z = -3.11$, $p = .02$) and near patches in condition (b) ($EMM_{cong} = .00$ versus $EMM_{incong} = .05$, $z = -5.22$, $p < .0001$). The effect of direction did not reach significance in the between-series conditions (c) ($EMM_{cong} = .29$ versus $EMM_{incong} = .20$, $z = 1.33$, $p = .56$) or (d) ($EMM_{cong} = .30$ versus $EMM_{incong} = .19$, $z = 1.69$, $p = .32$).

More importantly, the GLMM also provided evidence for the expected type × antecedent interaction, $\chi^2(1) = 12.01$, $p = .007$, displayed in Fig. 8.3. The visual impression from the figure is very much in line with a defective truth table pattern for the between-series condition. When moving to the "False" end of the scale—that is, to the right if the antecedent is a color as in condition (c), and to the left if the antecedent is a sphere as in condition (d)—the probability of an indeterminate response visibly increases. The antecedent slope differed significantly from 0 in condition (d), $b = -0.11$, 95% CI [−0.21, −0.02], although not in condition (c), $b = 0.04$, 95% CI [−0.06, 0.14]. In the within-series condition the slope in condition (a) also differed significantly from 0, $b = 0.15$, 95% CI [0.03, 0.28] (i.e., also tended to show the defective truth table pattern, albeit less pronounced, as Fig. 8.3 shows), but not in condition (b), $b = -0.01$, 95% CI [−0.14, 0.11]. In terms of differences between slopes, the slope in condition (d) was significantly steeper than the slope in condition (a), $z = 3.32$, $p = .005$. However, none of the remaining comparisons between slopes reached significance, $|z| < 2.2$, $p > .13$. Thus, responses in the between-series conditions tended to conform to the defective truth table pattern more than the responses in the within-series conditions (although not all relevant comparisons were significant). None of the other effects of the GLMM reached significance, largest $\chi^2(3) = 7.76$, smallest $p = .05$.

Taken together, these results show two things: (1) Guaranteed relevance provides an important cue for participants to override the if-heuristic which leads to an overall very low level of indeterminate responses. (2) In the absence of guaranteed relevance, the if-heuristic produces a defective truth table pattern, where the rate of indeterminate responses appears to be a function of antecedent position. In addition, the descriptive analysis slightly qualified the support for our hypothesis. When the strong relevance cue was absent, there was a marked increase in the rate of indeterminate responses, but also in the rate of "False" responses. This suggests that participants are somewhat split whether or not conditionals in the between-series condition were void or just false. This pattern is in line with recent research on the defective truth table (e.g., Schroyens, 2010). Another possibility is that the increased proportion of

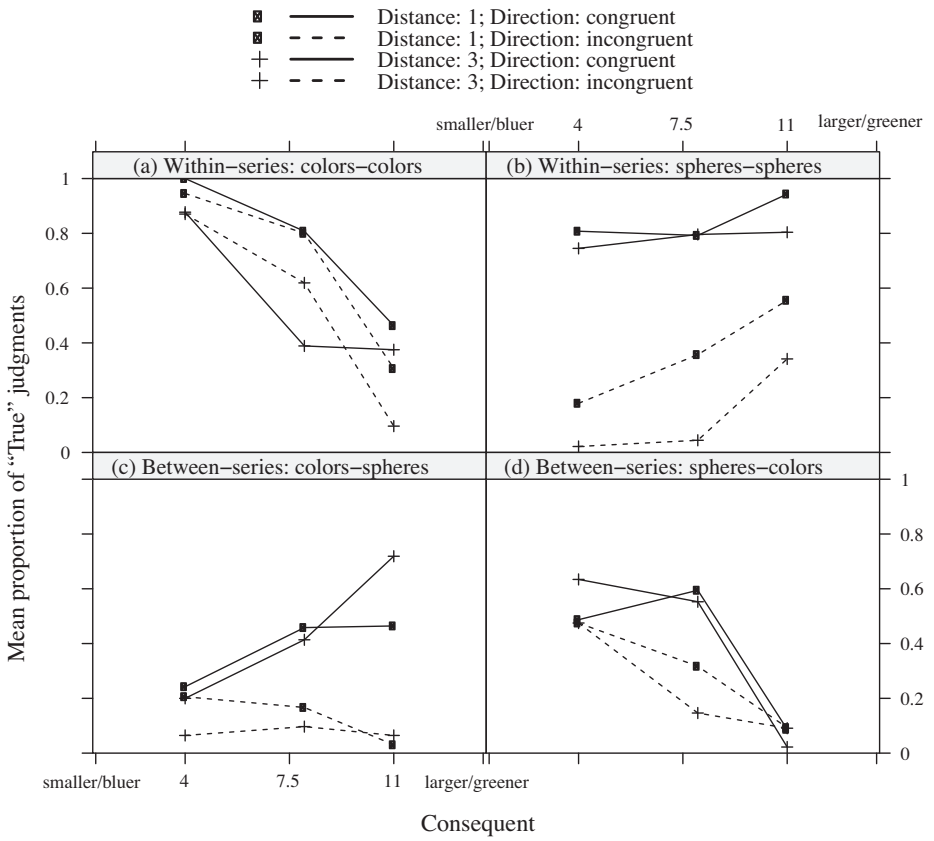

Fig. 8.4. Data from Experiment 3 relevant to HIT predictions concerning "True" versus "False" judgments (excluding all "Neither true nor false" responses) as a function of consequent (on the *x*-axis), type, distance between antecedent and consequent, and direction.

"False" responses is due to some participants in the between-series condition resorting to conjunctive responses, a pattern often found with abstract conditionals (e.g., Evans et al., 2003; Oberauer & Wilhelm, 2003).

8.3.2. *"True" versus "False" responses*

Mean response proportions of the "True" versus "False" judgments are displayed in Fig. 8.4. For the within-series conditions (top panels), the pattern appears to replicate those of Experiments 1 and 2. For condition (a), the pattern seems to replicate the results from Experiment 1 almost exactly with the exception of one data point: the middle point for incongruent patches with distance 3 seems too low. The pattern in condition (b) seems to match the results from Experiment 2 almost exactly.

For the between-series condition (lower panels), the pattern appears to be markedly different than for the within-series condition, with overall a considerably lower level of "True" judgments. Nevertheless, two main trends still appear to be present, to wit, larger rates of true judgments for congruent compared to incongruent conditionals, and visible consequent slopes.[12]

Model selection. As in the analyses of the previous experiments, we compared three GLMMs: one for HIT (i.e., fixed effects for direction, distance, consequent, type, and all interactions); a consequent model (i.e., consequent, type, and their interaction); and an antecedent–consequent model (i.e., antecedent, consequent, type, and their interactions). Based on the expected differences between the within-series and between-series condition, we performed this analysis separately for each condition. This decision made type a between-subjects factor with two levels in each condition. Results are displayed in Table 8.1 and show the expected difference between the two conditions. For the within-series condition, the HIT model provided, as expected, the best account, $\Delta AIC > 150$. For the between-series condition, the antecedent–consequent model provided the best account, $\Delta AIC > 78$.

Analysis of HIT model: Within-series condition. For the within-series conditions, our main prediction was a replication of the pattern found in the previous experiment. In line with this, we found support for the inferential strength hypothesis with both a main effect of direction, $\chi^2(1) = 25.49, p < .0001$, $OR > 2100$ ($EMM_{cong} = 1.00$ versus $EMM_{incong} = .61$), and a main effect of distance $\chi^2(1) = 42.95, p < .0001$, $OR > 24000$, ($EMM_1 = 1.00$ versus $EMM_3 = .32$). Again we did not find a direction × distance interaction, $\chi^2(1) = 0.00, p = .95$, but a type × direction interaction, $\chi^2(1) = 58.43, p < .0001$, and a type × direction × distance interaction, $\chi^2(1) = 12.22, p = .0005$. The last two effects indicate that the effect of direction and distance differed between the spheres–spheres

[12] Note again that congruency in the between-series condition is based on the position of the consequent.

Table 8.1
Model comparison of GLMMs on "True" versus "False" responses for Experiment 3.

Model	K_f	K_r	LL	AIC	ΔAIC	BIC	ΔBIC
			Within-series condition				
HIT	16	8/28	−403.20	910.41	0.00	1175.09	22.14
Consequent	4	2/1	−677.46	1368.91	458.50	1404.54	251.59
Antecedent–consequent	8	4/6	−512.66	1061.33	150.92	1152.95	0.00
			Between-series condition				
HIT	16	8/28	−341.51	787.01	78.69	1034.43	240.47
Consequent	4	2/1	−404.85	823.70	115.38	857.00	63.04
Antecedent–consequent	8	4/6	−336.16	708.32	0.00	793.97	0.00

Note. See Table 6.2. As in Experiment 2, the apparent better performance of the antecedent–consequent model compared to the HIT model for the within-series condition in terms of BIC is a consequence of the correlation parameters among the random effects. When removing those, the HIT model provides the best account in terms of both AIC, ΔAIC > 150, and BIC, ΔBIC > 95. For the between-series condition, removing the correlations does not affect the ordering of the models in terms of their AIC or BIC performance.

and colors–colors conditions (see supplemental materials for details).

We also again found support for the belief bias effect, a significant type × consequent interaction, $\chi^2(1) = 102.11$, $p < .0001$ ($b_{color} = -5.99$, 95% CI [−7.23, −4.75], versus $b_{spheres} = 1.86$, 95% CI [0.96, 2.76]). There was also evidence for the validity × belief analogue, a significant type × direction × consequent interaction, $\chi^2(1) = 11.51$, $p = .0007$. In the colors–colors condition, the effect of consequent was less pronounced for congruent ($b = -5.26$, 95% CI [−6.54, −3.98]) than for incongruent patches ($b = -6.72$, 95% CI [−8.22, −5.22]), in the spheres–spheres condition the effect of consequent was absent for congruent patches ($b = 0.55$, 95% CI [−0.29, 1.40]) but it was clearly there for incongruent patches ($b = 3.17$, 95% CI [1.83, 4.51]), and all slopes differed from each other ($p < .03$). We also found a main effect of type, $\chi^2(1) = 14.28$, $p = .0002$, OR = 650, indicating that the rate of true judgments was larger in the colors–colors condition (EMM = 1.00) than in the spheres–spheres condition (EMM = .74). None of the remaining interactions reached significance, largest $\chi^2(1) = 1.1$, smallest $p = .29$.

Analysis of HIT model: Between-series condition. For the between-series conditions, the results were in line with the visual inspection. We observed an effect of direction, $\chi^2(1) = 4.79$, $p = .03$, OR = 12.36, ($EMM_{cong} = .28$ versus $EMM_{incong} = .03$) and a type × consequent interaction, $\chi^2(1) = 20.41$, $p < .0001$ ($b_{colors—spheres} = 0.44$, 95% CI [0.10, 0.78], versus $b_{spheres—colors} = -0.56$, 95% CI [−0.85, −0.28]). In addition, we found a main effect of type, $\chi^2(1) = 8.77$, $p = .003$, OR = 0.13, indicating that the rate of true judgments was larger in the spheres–colors condition (EMM = .23) than in the colors–spheres condition (EMM = .04). None of the remaining effects reached significance, largest $\chi^2(1) = 3.2$, smallest $p = .07$.

In conclusion, in Experiment 3 we set out to test the *principle of relevant inference*, by pitting conditionals which refer to a single soritical series, whose relevance is guaranteed, against conditionals which refer to two soritical series, where relevance is suppressed. We predicted, and found, that when relevance was suppressed, the proportion of indeterminate responses was substantially and significantly higher, increasing as the antecedent rank became "falser," conforming to the defective conditional pattern. In contrast, when the antecedent and consequent were within a single series, we replicated the patterns found in Experiments 1 and 2.

9. Experiment 4

All previous experiments supported our *principle of relevant inference*, demonstrating that, under conditions of guaranteed relevance, the defective truth table pattern disappears. Moreover, Experiment 3 directly supported this principle by demonstrating that when this relevance was defeated, responses relapsed to the defective truth table. Thus far, we supported our *principle of bounded inference* by showing that factors hypothesized to affect the strength of inference from antecedent to consequent also affected truth evaluation, and that the evaluation pattern was subject to the same belief bias that affects other types of inference. Experiment 4 aimed to provide a more direct support for the principle of bounded inference, by using a measure of participants' subjective evaluations of bounded inference strength. Moreover, Experiment 4 is the most direct test of the core principle of HIT, the inferentialist principle that truth evaluations of conditionals are determined by the existence of an inferential connection between antecedent and consequent.

In this experiment, we returned to the stimuli set used in Experiment 1 (where belief bias was more in evidence), but added two direct measures: we asked participants to evaluate the strength of the inference from antecedent to consequent, as well as their metacognitive confidence in their response. The inference strength scale was borrowed from Elqayam et al. (2015), and is suitable for directly measuring the strength of informal inference. Participants are presented with the premise (in this case, the antecedent) and the conclusion (in this case, the consequent), and asked to rate the extent to which the conclusion follows from the premise on a scale from "Definitely does not follow" to "Definitely follows."

The metacognitive confidence scale was inspired by recent work on metacognition and reasoning, sometimes dubbed "meta-reasoning" (see, e.g., Ackerman & Thompson, 2015, 2017a, 2017b; Thompson, Prowse Turner, & Pennycook, 2011). This research domain, which branches off dual process theories, explores the psychological on/off switch for effortful, Type 2 processing. We will have more to say about metacognition and meta-reasoning in the General Discussion; for now we just note that meta-reasoning research inspired our work in two ways. First, it makes a firm distinction between direct judgments of inference such as validity or

strength (first-order measures), and confidence in those judgments (a second-order measure). Second, it provided the scale for metacognitive confidence, which we needed in order to estimate when the inference is considered to be satisficing, or strong enough. Recall that, according to HIT's principle of bounded inference, the inference from antecedent to consequent should be *strong enough* for a conditional to be evaluated as true. Consequently, we needed to ask for both inference strength judgments and metacognitive confidence, in order to capture the "strong" element as well as the "enough" element of this principle, respectively.

9.1. Predictions

We had four sets of predictions: first, we predicted replication of the pattern established in the three previous experiments, and especially Experiment 1. We also had three graded sets of predictions for the added variables that measure bounded inference, that is, inference strength and metacognitive confidence.

Replication. For the truth evaluation task, our predictions were essentially the same as in Experiment 1. We predicted a replication of the inferential strength effect, articulated as main effects of direction and distance. We left the distance × direction interaction again as an exploratory hypothesis. We also predicted a consequent effect (our belief bias analogue). As a minor prediction, we predicted a consequent × direction effect, a replication of the same effect from Experiments 1 and 2. As before, this prediction is less firm because the believability × validity interaction is not universal in belief bias.

Bounded inference. HIT's central thesis is that the mechanism that underlies the truth evaluation of conditionals is, by default, relevant, bounded inference from antecedent to consequent. In this experiment, bounded inference was measured by the twin parameters of inference strength judgment and metacognitive confidence. There are three graded interpretations of this thesis. If HIT is right, then truth evaluation should mimic the same response pattern as the variable measuring bounded inference, that is, inference strength and metacognitive confidence, providing a *basic* level of support for the thesis. Hence, we expected that inference strength judgment and metacognitive confidence would reflect the same pattern as truth evaluation. Specifically, we predicted inference strength judgment to be sensitive to distance and direction (the factors manipulating inferential strength) in the same way as truth evaluation; and we expected both inference strength judgment and metacognitive confidence to be sensitive to consequent effects, again in the same way as truth evaluation. We left any interaction effects between direction and distance as an exploratory hypothesis. As a minor prediction, we predicted a consequent × direction effect on inference strength judgment.[13]

To provide stronger support for HIT, our bounded inference variables (i.e., inference strength judgment and metacognitive confidence) should also *predict* truth evaluation, in a model in which truth evaluation is the criterion and all other variables—distance, direction, consequent, inference strength judgment, and metacognitive confidence—are predictors. The bounded inference variables should be significant predictors of truth evaluation, thus providing an *intermediate* level of support for HIT. At the *strongest* level of support, inference strength judgment and metacognitive confidence should be the *only* significant predictors in such a model, superseding all other predictors such as direction and distance, hence showing that bounded inference is the only explanation for truth evaluation.

9.2. Method

PARTICIPANTS

One hundred and thirty-three participants were recruited the same way as in the previous experiments. We used the same validation measures and exclusion criteria as in Experiment 2, which left us with 99 participants.[14] These participants spent on average 502 s on the experiment (SD: 186 s). Seventy-five of them had a university education and 24 had only a high school or secondary school education. Their mean age was 41 years (±11).

DESIGN AND MATERIALS

We used a trimmed-down version of the design from Experiment 1, with only the in-sight visual condition, a single named color condition (blue), and with two levels of distance, 1 and 3, manipulated entirely within participants. Each participant was presented with a total of 14 conditionals. Antecedent patches were in position 3, 6, 9, or 12, and for each antecedent patch, the consequents were +/−1 and +/−3 (with the exception of antecedent patches 3 and 12, for which −3 and, respectively, +3 were impossible).

Participants were asked to complete two tasks, presented in counterbalanced order, so that half of the participants completed the truth-evaluation task first (50 of the final participants), and half of the participants completed the bounded inference task first (49 of the final participants). The truth-evaluation task was identical to the one in Experiment 1, with the same soritical color series. Participants were first given a practice item, then presented with the set of 14 items, each on a separate page, in an individually randomized order.

The bounded inference task, presented separately, included two questions for each item: inference strength judgment and metacognitive confidence. The inference strength task presented participants with the antecedent of the conditional as a premise, and asked them to evaluate how strongly the consequent followed on a fully-labeled 7-point Likert-type scale: *Definitely does not follow*, *Follows very weakly*, *Follows weakly*, *Follows to some degree*, *Follows strongly*, *Follows very strongly*, and *Definitely follows*. This question

[13] We expected metacognitive confidence only to be affected by intuitive factors (in our task, consequent effects), since it is well-established in the metacognition literature that time-free confidence ratings ("Final Judgment of Confidence") are only sensitive to those factors (e.g., Thompson & Johnson, 2014; Thompson et al., 2011, 2013).

[14] In relation to Experiment 2, the $N = 99$ is consistent with the 2.5 times rule of Simonsohn (2015) for replication studies.

Fig. 9.1. Inference strength judgment and metacognitive confidence task example.

was followed by the metacognitive question directly underneath, taken from Thompson et al. (2011): "In providing my answer to the above question I felt: ...," which in turn was followed by a partially-labeled 7-point Likert-type scale with the labels *Guessing*, *Fairly certain*, and *Certain I'm right* in the extreme left, midpoint, and extreme right, respectively.[15] Following a separate practice item, the 14 test items were presented, each on a separate page, in an individually randomized order; see Fig. 9.1 for an example.

9.3. Results and discussion

A first analysis showed that the order in which participants worked on the two tasks did not affect the results (i.e., when including order in the models reported below, no effect involving order reached significance). Furthermore, for some GLMMs excluding order led to a *better* model fit than including order, indicating that models including order did not converge to the maximum likelihood estimates (for LMMs—see below—the difference in model fit was very small and not significant). Consequently, all results reported below are based on models without order as factor.

9.3.1. Indeterminate responses

Overall, participants classified 59.5 percent of the conditionals as true, 26.8 percent as false, and 13.6 percent as neither true nor false. In line with the previous results, only one of the 99 participant always responded with "Neither true nor false" while 38 participants never used this response (zero participants always responded with "True" and two never, and one participant always responded with "False" and eleven never). Fig. 9.2 displays the distribution of the individual response proportions; these are very similar to the results from Experiments 1 and 2 as well as from the inferential connection condition of Experiment 3.

9.3.2. "True" versus "False" responses

Fig. 9.3 displays the rate of "True" versus "False" responses. The pattern of results clearly resembles the pattern of results from Experiment 1, showing a strong consequent effect as well as an effect of distance. Only the effect of direction seems absent or at least strongly attenuated. We start this section with comparing the HIT model with competitor models before testing the presence of the predicted effects.

Model selection. Table 9.1 shows the GLMM model selection result for the three models corresponding to the main accounts. As before, the HIT model provides the best account, $\Delta AIC > 134$. This strongly indicates, once again, that only when considering all variables deemed relevant by HIT can a model provide an adequate account of "True" versus "False" judgments.

Analysis of HIT model. Regarding the effect of the inferential strength parameters—distance and direction—we did not see an effect of direction, $\chi^2(1) = 2.30, p = .13$.[16] But we found a main effect of distance again, $\chi^2(1) = 60.51, p < .0001$. Patches with distance 1 were more likely to be judged true ($EMM_1 = 1.00$) than patches with distance 3 ($EMM_3 = .49$), $OR > 590000$. In addition, we now found a direction × distance interaction, $\chi^2(1) = 9.73, p = .002$. However, this interaction only partly supported the predictions. There was no support for the predicted differential effect of distance within each level of direction. Instead, there was a strong effect of direction for patches with distance 1, $z = 2.14, p = .03$, $OR > 1000$ (although $EMM_{cong} = EMM_{incong} = 1.00$), but no

[15] Note that this is not the full Thompson two-response paradigm, in which participants first provide a quick response, which is then followed by a second response with more time for effortful processing; we only adopted the confidence question as a measure of satisficing. We will return to this issue in the General Discussion.

[16] Just as for Experiment 2, we were unable to reliably obtain *p*-values for the model including correlations among random slopes. Consequently, this section is based on a model without correlations among random slopes. See Table 4 in the supplemental materials for full results.

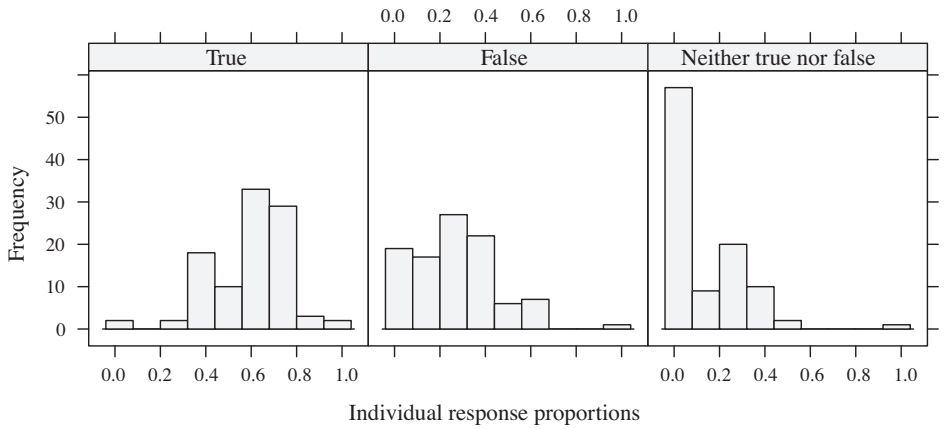

Fig. 9.2. Histogram of individual response proportions of Experiment 4.

effect of direction for patches with distance 3, $z = 0.08$, $p = .94$, OR = 1.20 ($EMM_{cong} = .51$ and $EMM_{incong} = .47$).

In support of our belief bias hypothesis, we found a strong main effect of consequent, $\chi^2(1) = 112.17$, $p < .0001$, $b = -5.61$, 95% CI $[-7.14, -4.08]$, replicating the pattern observed in Experiments 1 and 2 and in the within-series condition of Experiment 3. We also found a consequent × distance interaction, $\chi^2(1) = 7.20$, $p = .007$ (supplemental materials Fig. 10). Consequent effects were strongest (i.e., largest absolute value) for patches with distance 3, $b = -6.61$, 95% CI $[-8.51, -4.71]$, and less strong (i.e., smaller absolute value) for patches with distance 1, $b = -4.61$, 95% CI $[-6.05, -3.18]$.

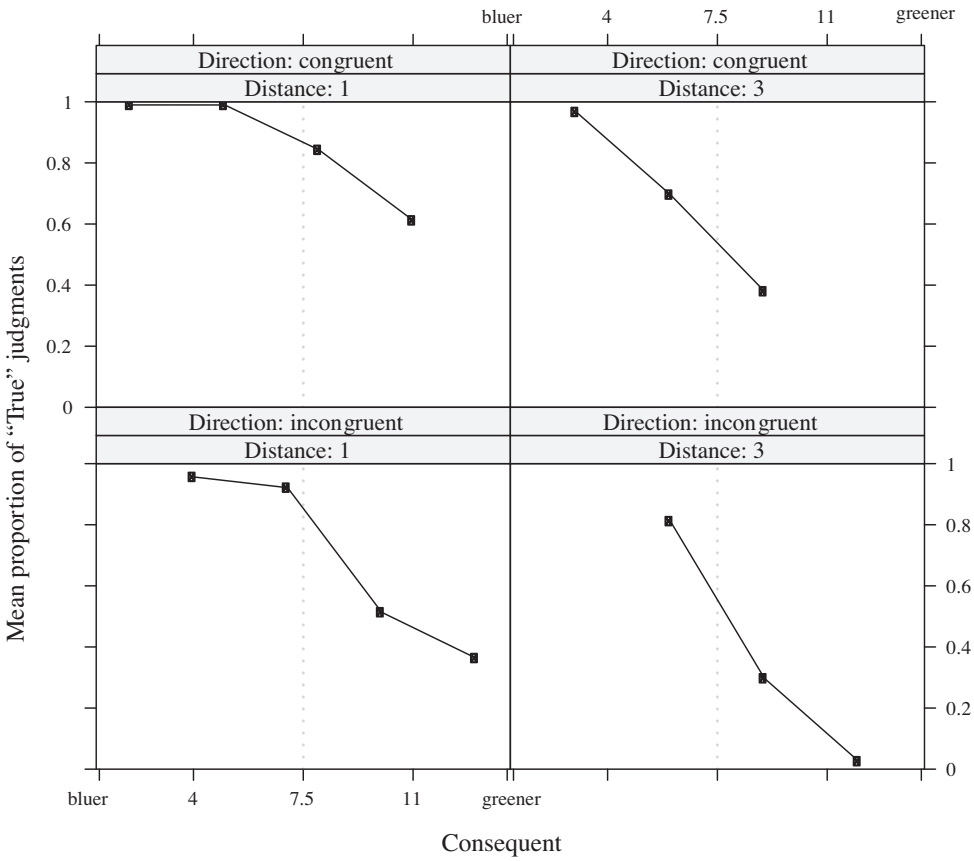

Fig. 9.3. Data from Experiment 4 relevant to HIT predictions concerning "True" versus "False" judgments (excluding all "Neither true nor false" responses) as a function of consequent (on the *x*-axis), distance between antecedent and consequent, and direction. (Fig. 9 in the supplemental materials is a version of this figure that includes the "Neither true nor false" responses.)

Table 9.1
Model comparison of GLMMs on "True" versus "False" responses for Experiment 4.

Model	K_f	K_r	LL	AIC	ΔAIC	BIC	ΔBIC
HIT	8	8/28	−311.59	711.17	0.00	934.99	18.53
Consequent	2	2/1	−492.28	994.56	283.38	1019.99	103.53
Antecedent–consequent	4	4/6	−408.62	845.24	134.07	916.46	0.00

Note. See Table 6.2. As for Experiment 2, the apparent better performance of the antecedent–consequent model in terms of BIC is solely an effect of the correlation among random slopes. After removing those, HIT provides the best account in terms of both AIC, ΔAIC > 162, and BIC, ΔBIC > 121.

9.3.3. Inference strength measure

For the inference strength measure we predicted the same pattern of results as for the truth judgments. Fig. 9.4 depicts the data and reveals a very similar pattern as the one seen in Fig. 9.3.[17] One can clearly see an effect of consequent which appears stronger for patches with distance 3 than for patches with distance 1. In addition to this, judgments of inference strength appear to be stronger for congruent patches (top row) than for incongruent patches (bottom row). Further inspection also suggests a similar nonlinear relationship. In the GLMM, this nonlinearity in the effect of consequent was captured by the logistic linking function, but this solution is not possible for linear mixed models (LMMs; Baayen, Davidson, & Bates, 2008). We therefore tried to capture the nonlinearity with a simple quadratic effect of consequent in addition to the linear effect of consequent. As the comparison of observed and predicted data in Figs. 9.4 and 9.5 shows, this was sufficient to adequately capture the trends in the data.

We entered the individual responses to the inference strength question to an LMM with fixed effects for direction, distance, consequent (linear and quadratic), as well as their interactions. The results showed that all effects, with the exception of the three-way interaction of direction, distance, and the quadratic component of consequent ($\chi^2(1) = 0.99, p = .32$), reached significance,

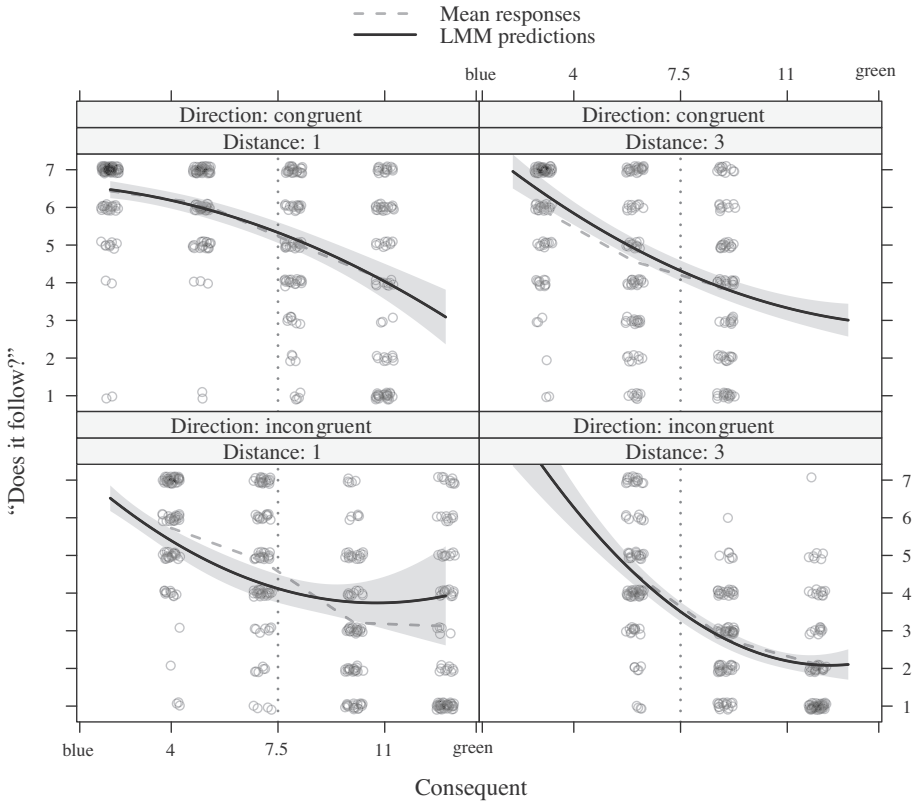

Fig. 9.4. Responses to the inference strength measure and corresponding LMM predictions as a function of variables relevant to HIT. Individual data points are plotted with 70 percent transparency and with jitter added. Gray areas indicate 95 percent confidence bands.

[17] For this and the analysis of metacognitive judgments, we included all trials, even those for which participants decided that the conditional was neither true nor false. We did not see any reason to exclude trials here. However, as suggested by an anonymous reviewer, excluding the indeterminate responses makes the analysis more similar to the one reported for the "True" versus "False" judgments. Consequently, we repeated both analyses excluding the indeterminate responses. This led to the same pattern of significant and non-significant results as reported here.

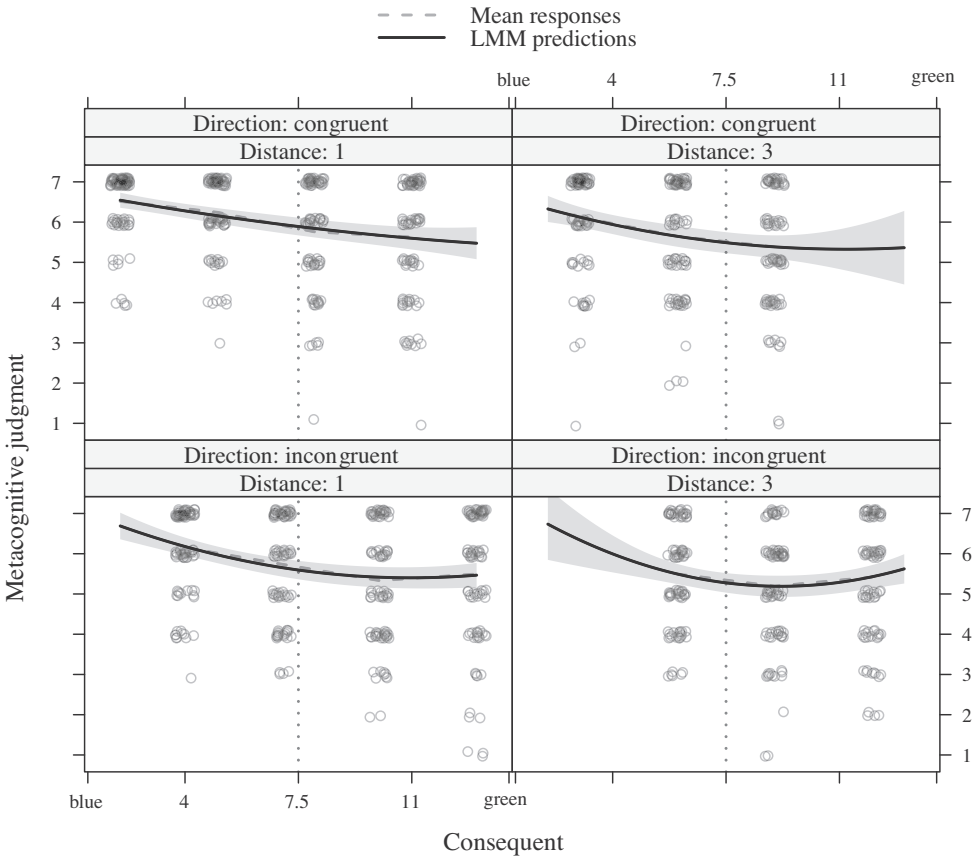

Fig. 9.5. Responses to the metacognitive confidence measure and corresponding LMM predictions. NB: Mean responses (in the dashed lines) are hardly visible because they are so well-aligned with the predictions (in the solid black lines).

smallest $\chi^2(1) = 5.08$, largest $p = .02$. See supplemental materials Table 5 for full results. The corresponding model predictions are plotted in Fig. 9.4 as black lines on top of the data points.

We also replicated the truth judgment pattern for inference strength judgments, thus supporting our inferential strength hypothesis: congruent patches received stronger inference strength ratings than incongruent patches, EMM = 4.73 versus EMM = 3.91, and patches with distance 1 received stronger inference strength ratings than patches with distance 3, EMM = 4.83 versus EMM = 3.82. Furthermore, we replicated the belief bias effect and found a clear linear effect of consequent, $b = -0.37$, 95% CI [−0.43, −0.31]. In line with the visual impression, we found a quadratic effect of consequent, $b = 0.025$, 95% CI [0.016, 0.034].

Finally, we replicated the analogue to the believability × validity interaction: the (linear) effect of consequent was strongest for incongruent patches with distance 3, $b = -0.58$, 95% CI [−0.68, −0.48], which differed from all other effects of consequent, smallest $z = 4.6$, all $ps < .0001$. The next strongest effect was for incongruent patches with distance 1, $b = -0.36$, 95% CI [−0.42, −0.30], which differed from the effect for congruent patches with distance 3, $b = -0.24$, 95% CI [−0.34, −0.13], $z = -2.57$, $p = .03$, but not from the effect for congruent patches with distance 1, $b = -0.31$, 95% CI [−0.37, −0.25], $z = 1.93$, $p = .11$. The two effects for congruent patches did not differ from each other, $z = -1.53$, $p = .13$.

9.3.4. Metacognitive confidence

For the metacognitive confidence judgments we only predicted effects of consequent. Fig. 9.5 depicts the data; an eyeball test suggests somewhat weaker effects compared to those seen in Figs. 9.3 and 9.4, although the effect of consequent still seems strong. To statistically assess this pattern, we followed the same approach as for the inference strength measure and estimated an LMM with the metacognitive judgments as dependent variable and fixed effects for direction, distance, consequent (linear and quadratic), as well as their interactions (see supplemental materials Table 6 for full results). This analysis again supported our bounded inference prediction at the basic level, showing a parallel pattern to that of truth evaluation. It revealed the predicted belief bias effect, a main effect of consequent (linear component), $b = -0.10$, 95% CI [−0.13, −0.07], $\chi^2(1) = 31.16$, $p < .0001$. In addition, we found a main effect of the quadratic component of consequent, $b = 0.015$, 95% CI [0.005, 0.025], $\chi^2(1) = 8.71$, $p = .003$.

We also found evidence for an effect of inferential strength on metacognitive judgments: a main effect of direction, $\chi^2(1) = 8.57$, $p = .003$, indicating that congruent patches received stronger certainty ratings, EMM = 5.69, than incogruent patches, EMM = 5.43, as well as a main effect of distance, $\chi^2(1) = 21.96$, $p < .0001$, indicating that consequent patches with distance 1,

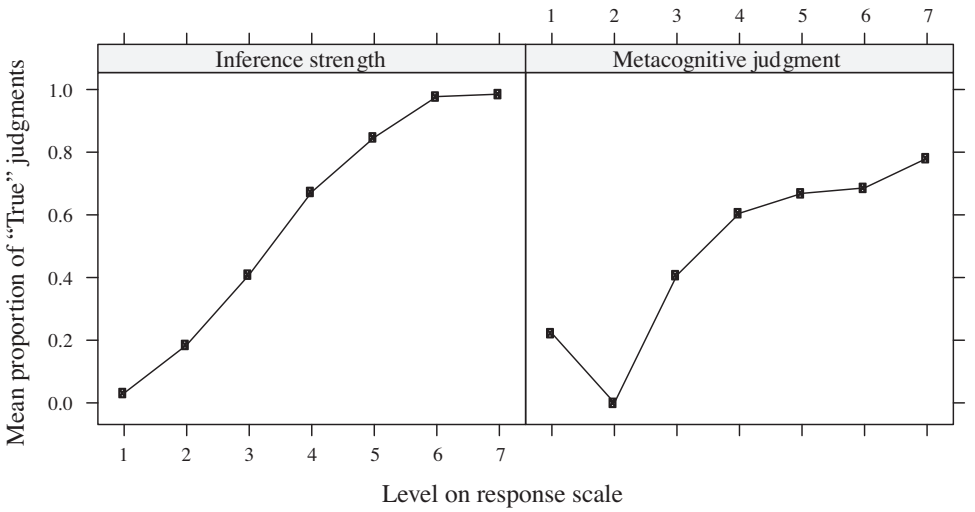

Fig. 9.6. "True" versus "False" judgments (excluding all "Neither true nor false" responses) as a function of inference strength judgment and metacognitive confidence.

$EMM_1 = 5.74$, received stronger certainty ratings than congruent patches with distance 3, $EMM_3 = 5.38$. Finally, we found a weak interaction of direction with the quadratic component of consequent, $\chi^2(1) = 4.46$, $p = .03$, indicating that the quadratic effect is absent for congruent patches, $b = 0.008$, 95% CI [−0.003, 0.018], but stronger for incongruent ones, $b = 0.023$, 95% CI [0.013, 0.033]. As in the truth judgments and inference strength measure, this latter interaction can be interpreted as a validity × believability analogue.

9.3.5. Combined model

Recall that in Section 9.1 we introduced three possible levels of support for our bounded inference hypothesis: basic, in which the bounded inference variables, inference strength and metacognitive confidence, mimic the pattern found for truth evaluations; intermediate, in which these bounded inference variables also significantly predict the truth evaluation pattern; and the strongest level, in which they do so exclusively. To test if the data supported our bounded inference prediction at the intermediate or even strongest level, we tested if inference strength judgment and metacognitive confidence predict truth judgments. Fig. 9.6 shows those relationships. For both variables an increase is strongly associated with a higher probability of responding with "True." To test if these effects can explain the effects of the variables deemed relevant by HIT (in a statistical sense), we combined each of the two variables separately with the HIT model for truth judgments (a joint model with both variables did not converge). In contrast to the other tests reported in this article, we were unable to employ likelihood ratio tests and had to resort to Wald tests instead (Fox, 2008).

Inference strength. We estimated a GLMM with truth judgments as dependent variable and fixed effects for direction, distance, consequent, and their interactions plus a fixed effect for inference strength (after centering at the midpoint of the scale). As expected, we found a strong effect of inference strength judgment, $b = 9.87$, 95% CI [4.80, 14.95], $\chi^2(1) = 14.54$, $p = .0001$, thus providing support for our bounded inference hypothesis at the intermediate level. Inference strength judgment could however not explain all effects in the data, failing to provide full support for our bounded inference hypothesis at the strongest level. However, the effects were much attenuated, giving partial support to this level of the hypothesis. We did find an effect of distance, $\chi^2(1) = 3.93$, $p = .05$, $EMM_1 = 1.00$ versus $EMM_3 = 0.79$, $OR > 1000$, but the consequent effect was weaker, $b = -4.28$, 95% CI [−7.84, −0.73], $\chi^2(1) = 5.58$, $p = .02$. Moreover, none of the other effects reached significance, largest $\chi^2(1) = 1.35$, smallest $p = .25$ (see Table 7 in the supplementary materials for full results).

Metacognitive confidence. We estimated another GLMM on the truth judgments, but this time we added a fixed effect for metacognitive judgment (centered). This analysis showed the expected effect of metacognitive judgments, $b = 1.07$, 95% CI [0.23, 1.91], $\chi^2(1) = 6.25$, $p = .01$, again providing support for our bounded inference hypothesis at the intermediate level. However, this effect was unable to fully explain the other effects. We still found effects of consequent, $b = -4.42$, 95% CI [−6.32, −2.52], $\chi^2(1) = 20.83$, $p < .0001$, distance $\chi^2(1) = 10.16$, $p = .001$, $EMM_1 = 1.00$ versus $EMM_3 = 0.57$, $OR > 1000$, and a direction × distance interaction, $\chi^2(1) = 7.79$, $p = .005$. In addition we now also found a three-way interaction of consequent × direction × distance, $\chi^2(1) = 5.29$, $p = .02$ (see Fig. 11 in the supplementary materials for a depiction of this interaction and Table 8 for full results). Thus, our bounded inference hypothesis was only partly supported at the strongest level.

In sum, the pattern of truth evaluations in Experiment 4 broadly replicated the one found in Experiment 1, with a low prevalence of indeterminate responses supporting our if-heuristic override hypothesis. Although there was no main effect of direction, we did find the usual distance effects, as well as a direction × distance interaction, with some support for the inferential strength hypothesis. We also supported the belief bias hypothesis via an effect of believability (consequent) as well as a validity × belief (direction × consequent) interaction. Furthermore, we tested a novel hypothesis unique to this experiment, the bounded inference hypothesis, and fully supported it at the basic and the intermediate level, and partly at the strongest level. Judgments of inference strength and metacognitive confidence displayed the same pattern as truth evaluations; and combined models showed that each of these two

additional measures were strong and significant predictors of truth evaluation, and that some of the other predictors were attenuated (although this worked better for inference strength judgment than for metacognitive confidence).

10. General discussion

In this paper, we presented HIT, a new theory of conditionals drawing on philosophical and psychological insights. According to this theory, people judge the truth value of a conditional by assessing the strength of the inferential connection between its antecedent and consequent, where the inferential connection may consist of any combination of deductive, inductive, and abductive steps. At the processing level, HIT postulates a dual processing framework, in which both intuitive, resource-frugal processes (Type 1) and effortful processes (Type 2) play a role. We proposed that people construct a relevant mental representation, which by default is the one in which there is an inferential relation between antecedent and consequent (*principle of relevant inference*). This relation need only be strong enough, in the sense of being subjectively supported (*principle of bounded inference*). People judge the conditional as true when there is a strong enough inferential connection between antecedent and consequent; false when the connection is weak or there is an argument from the antecedent to the negation of the consequent; and neither true nor false when there is no inferential connection at all, that is, when relevance cues fail. Because false antecedent cases tend to fall in the third category, HIT is able to reconcile our expectation of an inferential relation between antecedent and consequent with the accumulated psychological evidence for the defective truth table.

To test HIT, we designed a novel experimental task, the *soritical truth table task*, presenting it to a total of 893 participants across four experiments. The task was specifically designed to guarantee relevance of the consequent to the antecedent even when the antecedent is false, thus overriding the if-heuristic (the attentional cue which focuses speakers' attention on the antecedent being true), and circumventing the defective truth table response pattern. We varied parameters pertaining to inference strength such as distance between the stimuli and the direction of inference. Participants were given the usual three response options, "True," "False," and "Neither true nor false." Additionally, in Experiment 3 we directly manipulated relevance, using a variation of the task in which participants needed to evaluate conditionals within- versus between-soritical series. Experiment 4 also measured inference strength and metacognitive confidence, designed to tap directly into the strength of the bounded inference from antecedent to consequent. Table 10.1 sums up our main findings across the four experiments.

HIT predicts a unique response pattern, much of which was strongly supported by our data. First, since the soritical truth table task guarantees relevance, we predicted a massive majority of determinate ("True," "False") responses. That was also what we found, which supported our *if-heuristic override* hypothesis. Only a minority of the responses, and, even more importantly, a very small percentage of the responses to conditionals with false antecedents, were indeterminate—this, in stark contrast to the usual findings in classical truth table tasks, the defective truth table, in which indeterminate responses to false antecedent cases tend to be prevalent. We also found that when the task we designed undermined relevance (in Experiment 3), participants reverted to the defective truth table pattern. Furthermore, our results supported the *inferential strength* hypothesis, with effects of distance and direction in all four experiments.

Perhaps our most striking finding is the belief bias analogue which we predicted and which the data strongly supported. If conditionals are indeed inferential, they should display the same pattern found to hold for almost any type of inference, be it deductive or non-deductive; that is, they should display belief bias, the tendency to judge an inference based on the believability of the conclusion. Accordingly, we predicted a main effect of the believability of the consequent, analogous to the main effect of conclusion believability in classic belief bias. This effect replicated strongly and consistently across all four experiments. We also

Table 10.1
Summary of main results from Experiments 1–4.

Hypothesis	*Effects*	*HIT prediction*	*Experiment*					
			1	2	3R	4A	4B	4C
If-heuristic override	*Indeterminate responses*	Small proportion of indeterminate responses	✔	✔	✔	✔	NA	NA
Inference strength	*Direction*	Congruent > incongruent	✔	✔	✔	✗	✔	✔
	Distance	Near > far	✔	✔	✔	✔	✔	✔
Belief bias	*Consequent (belief bias 1)*	True > False	✔	✔	✔	✔	✔[b]	✔
	Consequent × validity (belief bias 2)	Stronger effect of consequent when direction and distance provide no reliable cues	✔[c]	✔[d]	✔[e]	✔[f]	✔[g]	✔[h]

Notes. Experiments 1, 2 and 3 measured truth evaluation. Predictions for Experiment 3 are only for the within-series condition where relevance is guaranteed (indicated here as 3R). Experiment 4 measured truth evaluation (4A), judgment of inference strength (4B), and metacognitive confidence (4C). Green checkmarks indicate support from the data, the red cross indicates lack of support; *a*: effect of direction for distance 1, but no effect of direction for distance 3; *b*: linear and quadratic; *c*: consequent × direction × color; *d*: consequent × direction; consequent × direction × distance; *e*: type × direction × consequent; *f*: consequent × distance; *g*: consequent × direction × distance; *h*: consequent × direction (quadratic).

predicted, and observed, an interaction effect between consequent position and factors relating to inferential strength, viz., distance, direction, or both. This effect is analogous to the validity × believability interaction in syllogistic belief bias. These results provided conceptual replication across the four experiments, although with some variations, a variability consistent with the extant literature on belief bias, in which the believability × validity effect of belief bias is not quite as robust as the main effect of believability.

Lastly and importantly, in Experiment 4 we directly tested the inferentialist core idea of HIT, according to which inferential connections determine truth evaluations of conditionals. We also tested the idea of bounded inference—that these connections need only be strong enough in the sense of being subjectively supported. We postulated three levels of support for the principle of bounded inference (the results for the relevant variables are depicted in Table 10.1): basic, in which the pattern for the bounded inference variables mimics that of truth evaluation; intermediate, in which bounded inference variables predict truth evaluation; and strongest, in which they do so exclusively. We found full support for the basic and intermediate levels, and partial support for the strongest level: the effects of other predictors attenuated when inference strength, but not when metacognitive confidence, was in the model. This may be due to the fact that we were unable to test both inference strength and metacognitive confidence in the same model. Another possible explanation is that we have not used the full meta-reasoning paradigm developed by Thompson and colleagues (e.g., Ackerman & Thompson, 2015, 2017a, 2017b; Thompson et al., 2011). Perhaps a *combination* of metacognitive measures would hit closer to the mark. We leave such exploration to future work.

In conclusion, our findings provide extensive, robust, and consistent support for HIT. The strength of inference from antecedent to consequent clearly has a major role to play in how we evaluate the truth of conditionals. We do not claim that this is *all* there is to evaluating conditionals: recall that some of the effects in Experiment 2 were attenuated, and that some predictors in Experiment 4 still played a role even when inference strength judgment or metacognitive confidence were in the model. We cannot entirely rule out factors beside inferential connections, but what we can say with confidence is that such factors cannot explain away our findings in any significant way. In particular, the pervasive and strong belief bias analogue, which can only be predicted and explained by a theory which regards conditionals as subject to inferential connections, provides strong psychological validation of our theory.

10.1. HIT, the Equation, and the New Paradigm

We are by no means the first to suggest that natural language conditionals embody a relationship between antecedent and consequent. We have already reviewed HIT's predecessor in the philosophical literature, inferentialism as formalized by Krzyżanowska et al. (2014). Within psychology, related ideas include Cheng and Holyoak's (1985) pragmatic reasoning schemas, one of which features causal-temporal relations between antecedent and consequent. Some of the work in mental model theory on spatial and temporal relations within conditionals touches on the idea that the antecedent should be related to the consequent (e.g., Juhos, Quelhas, & Johnson-Laird, 2012). Moreover, the idea takes center stage in theories of causal conditionals, especially those that employ causal Bayes nets (Hall, Ali, Chater, & Oaksford, 2016; Oaksford & Chater, 2013, 2014). Ali, Chater, and Oaksford (2011) refer approvingly to the position of Barwise and Perry (1983) that causal relations are at the semantic core of the conditional, and their results give some support to this view (see also Fernbach & Erb, 2013). Such theories strongly endorse the idea of inferential connections between antecedent and consequent, especially in the context of causal relations. How exactly inference of causal connections between antecedent and consequent fits into the larger picture of inferential connection is another important question still awaiting future work. Our soritical truth table task draws on abstract materials, in which the inferential connection is non-causal; it would be interesting to compare this to performance on an analogous causal task.

One possible interpretation is that causal Bayes nets can explain distance effects as a transitive chain from patch to adjacent patch, in which participants first infer from, say, patch number 1 to patch number 2, and then from patch number 2 to patch number 3.[18] Such chains may be driven by causality, but they may also be driven by inference. This (non-causal) interpretation of causal Bayes nets makes it even closer to HIT; indeed, HIT might be considered as much a specific articulation of causal Bayes nets as it is of the suppositional conditional.

Causal Bayes nets also make a good starting point for discussing the role of HIT within the broader framework of the New Paradigm in psychology of reasoning. We see HIT as falling squarely within the New Paradigm, as HIT has a natural affinity to theories of conditionals within this family, such as the psychological suppositional conditional and causal Bayes nets analyses of conditionals. The informal type of inference postulated by HIT, in which inferential connections only need be strong enough, also fits nicely with work on informal inference within the New Paradigm (e.g., Hahn & Oaksford, 2007; Mercier & Sperber, 2011).

At this stage of theoretical development, we do not have an over-arching formal theory (Marr's computational level of analysis) to model the subjective strength of inferential connections within conditionals. Hahn and Oaksford's Bayesian model of informal inference is a good candidate, but by no means the only one. It is entirely possible that different models are needed to account for inferential connections within different types of conditionals. We relegate to future work an investigation of this possibility. A further goal will be to link HIT to approaches (e.g., the dual-source model of Singmann, Klauer, & Beller, 2016) that aim to model the inference process underlying arguments involving conditionals.

Similarly, it is an open question what HIT's inferential principles imply vis-à-vis the Equation. Recent evidence suggests that the Equation does not hold for conditionals whose antecedent is not positively probabilistically relevant to their consequent (Skovgaard-Olsen et al., 2016). Whether and to what extent the Equation holds under different types of inferential connections still needs to be explored. We flag this here as another avenue for future research, though see Douven (2017b) for some first thoughts on what

[18] We thank Mike Oaksford for this suggestion.

inferentialists might want to say about the Equation.

On the processing side (Marr's algorithmic level of analysis), HIT takes as a departure point Hypothetical Thinking Theory. This dual processing model is based on a default-interventionist approach, rather than a parallel-competitive one (Evans, 2007b). The main difference between these models of dual processing is the question of whether Type 1 processes precede Type 2 processes (default-interventionist), or both types of processes proceed in tandem from the start (parallel-competitive). The evidence is equivocal and fraught with debate (see, e.g., Handley, Newstead, & Trippas, 2011), for example over the question of whether intuitions about logic and probability exist and, if they do exist, of how to interpret them (for the debate concerning logical intuitions see, e.g., De Neys, 2012; and cf. Klauer & Singmann, 2013). Going into the debate in any detail is beyond the scope of this work. For now we just note that, although HIT is compatible with the default-interventionist model, it would be easy enough to fit it with the parallel-competitive model. For example, the notion of default processes can be replaced with its closely related counterpart from the parallel-competitive model, the concept of fast, shallow Type 2 processing (De Neys & Glumicic, 2008). Similarly, the idea of intuitive logic can be easily subsumed—and even further developed—under HIT's suggestion that, by default, people interpret conditionals as postulating an inferential connection between their antecedent and consequent, where that connection may involve (but is not limited to) deductive, inductive, or abductive inference (or any combination of these).

This is also the place to note that, even within the New Paradigm, there is some debate over the if-heuristic and how explanatory it can be considered to be (e.g., Oaksford & Stenning, 1992). One potential explanation of our findings is that the contrast class in our paradigm is restricted to the other 13 stimuli in the series (be they color patches or spheres), whereas in everyday conditionals this is not so obviously the case.[19] However, note that the contrast class in Experiment 3's between-series conditions was also restricted (albeit to 27 other stimuli rather than 13), yet relevance was still suppressed. We defer further exploration of this issue to future work.

The design of Experiment 4 was partly inspired by another recent development within the New Paradigm, research in meta-cognition and meta-reasoning (e.g., Ackerman & Thompson, 2015, 2017a, 2017b; Thompson et al., 2011, 2013). The departure point for meta-reasoning research is that Type 2 processes require cognitive effort, whereas people tend to be cognitively lazy. The question then arises what regulatory mechanism triggers this extra cognitive effort. Meta-reasoning research focuses mainly on "Feeling of Rightness," or FOR—the metacognitive experience whose function is to signal when additional cognitive resources are necessary. FOR is usually measured by asking participants how confident they are in their responses. This was the measure that Experiment 3 borrowed from this line of research. Meta-reasoning research, however, usually employs the "two-response paradigm": participants are asked to provide an initial fast response, followed by confidence rating, and then a slower, more considered response to the same task, again followed by confidence rating. The term FOR is reserved for the first confidence rating, whereas the final confidence rating is termed "Final Judgment of Confidence", or FJC (e.g., Thompson et al., 2011, 2013).

We should clarify that, although we used a metacognitive *measure*, Experiment 4 did not aim to explore a metacognitive *research question*. Metacognitive research is about regulatory processes, whereas our research question focused on the diverse inferential connections that may exist between a conditional's antecedent and consequent, and on the nature of those connections. Meta-reasoning inspired our work, providing a convenient measure, which together with inference strength captured the essence of bounded inference. But we did not employ the two-response paradigm, which was beside the point for the purposes of the current study. Accordingly, we avoided using the terms associated with this paradigm (FOR, FJC), referring instead to "metacognitive confidence" simpliciter. The nature of the metacognitive processes underlying HIT is a pertinent research question. We have not addressed it at this stage as it was beyond the scope of this study; it still awaits future work. Potential studies in this vein must draw on the two-response paradigm, where both FOR and FJC can provide valuable clues to the nature of the processes underlying the appreciation of inferential connections within conditionals. Another potentially useful angle, strongly related to satisficing, is the idea of a "stopping rule" (Ackerman, 2014): the metacognitive mechanism which determines when the goal of the inference has been attained.

A good theory of conditionals is the Holy Grail for the psychology of reasoning as well as for philosophical logic. As two minimal desiderata, such a theory should cover both empirical evidence and semantic intuition. With this paper, we have taken some first steps toward incorporating a hitherto neglected semantic intuition into a psychological theory of conditionals.

Acknowledgments

We thank Rakefet Ackerman, John Anderson, Ruth Byrne, Jonathan Evans, Ulrike Hahn, Keith Holyoak, Art Markman, Mike Oaksford, as well as two anonymous referees for helpful comments on a previous version of this paper. We are also grateful to audiences in London, Munich, Paris, Providence, and Vienna, for stimulating questions and remarks. Our greatest debt is to David Over, for extensive comments on the design and on several earlier versions of this paper, and for many helpful discussions on conditionals and related matters.

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General, 143*, 1349–1368.
Ackerman, R., & Thompson, V. A. (2015). Meta-reasoning: What we can learn from meta-memory. In A. Feeney, & V. A. Thompson (Eds.). *Reasoning as memory* (pp. 164–178). Hove UK: Psychology Press.
Ackerman, R., & Thompson, V. A. (2017a). Meta-reasoning: Shedding meta-cognitive light on reasoning research. In L. J. Ball, & V. A. Thompson (Eds.). *International*

[19] We thank Mike Oaksford for this suggestion.

handbook of thinking and reasoning (pp. 1–15). Oxon UK: Routledge.
Ackerman, R., & Thompson, V. A. (2017b). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Science, 21*, 607–617.
Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional reasoning: Mental models or causal models. *Cognition, 119*, 403–419.
Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*, 527–535.
Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.
Baratgin, J., Douven, I., Evans, J. St. B. T., Oaksford, M., Over, D. E., & Politzer, G. (2015). The new paradigm and mental models. *Trends in Cognitive Sciences, 19*, 547–548.
Baratgin, J., Over, D. E., & Politzer, G. (2013). Uncertainty and the de Finetti tables. *Thinking & Reasoning, 19*, 308–328.
Barrouillet, P., Gauffroy, C., & Lecas, J.-F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review, 115*, 760–771.
Barwise, J., & Perry, J. (1983). *Situations and attitudes.* Cambridge MA: MIT Press.
Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. Retrieved from <http://arxiv.org/abs/1506.04967>.
Beller, S., Bender, A., & Kuhnmünch, G. (2005). Understanding conditional promises and threats. *Thinking & Reasoning, 11*, 209–238.
Bennett, J. (2003). *A philosophical guide to conditionals.* Oxford: Oxford University Press.
Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1–21.
Braine, M. D. S., & O'Brien, D. P. (1991). A theory of if: Lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182–203.
Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391–416.
Corner, A., Hahn, U., & Oaksford, M. (2011). The psychological mechanism of the slippery slope argument. *Journal of Memory and Language, 64*, 133–152.
Cruz, N., Over, D. E., Oaksford, M., & Baratgin, J. (2016). Centering and the meaning of conditionals. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.). *Proceedings of the 38th annual conference of the cognitive science society* (pp. 1104–1109). Austin TX: Cognitive Science Society.
Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23*, 646–658.
Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition, 19*, 274–282.
de Finetti, B. (1995). The logic of probability. *Philosophical Studies, 77*, 181–190.
De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7*, 28–38.
De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248–1299.
Douven, I. (2013). Inference to the best explanation, Dutch books, and inaccuracy minimisation. *Philosophical Quarterly, 63*, 428–444.
Douven, I. (2016a). Explanation, updating, and accuracy. *Journal of Cognitive Psychology, 28*, 1004–1012.
Douven, I. (2016b). *The epistemology of indicative conditionals.* Cambridge: Cambridge University Press.
Douven, I. (2017a). Abduction. In Zalta E. N. (Ed.), *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/sum2017/entries/abduction/>.
Douven, I. (2017b). How to account for the oddness of missing-link conditionals. *Synthese, 194*, 1541–1554.
Douven, I. (2017c). Inference to the best explanation: What is it? And why should we care? In T. Poston, & K. McCain (Eds.). *Best explanations: New essays on inference to the best explanation* (pp. 4–22). Oxford: Oxford University Press.
Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2017). Conditionals and inferential connections: A comparison of semantic and psychological models. Manuscript.
Douven, I., & Krzyżanowska, K.H. (2018). The semantics–pragmatics interface: An empirical investigation. In Capone, A. (Ed.), *Further advances in pragmatics and philosophy* (Vol. II). New York: Springer (in press).
Douven, I., & Mirabile, P. (2018). Best, second-best, and good-enough explanations: How they matter to reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* (in press).
Douven, I., & Schupbach, J. N. (2015). The role of explanatory considerations in updating. *Cognition, 142*, 299–311.
Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition, 117*, 302–318.
Douven, I., & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science, 37*, 711–730.
Elqayam, S. (2017). The new paradigm in psychology of reasoning. In L. J. Ball, & V. A. Thompson (Eds.). *International handbook of thinking and reasoning* (pp. 130–150). Oxon, UK: Routledge.
Elqayam, S., & Evans, J. St. B. T. (2013). Rationality in the new paradigm: Strict versus soft Bayesian approaches. *Thinking & Reasoning, 19*, 453–470.
Elqayam, S., & Over, D. E. (2012). Probabilities, beliefs, and dual processing: The paradigm shift in the psychology of reasoning. *Mind & Society, 11*, 27–40.
Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning. *Thinking & Reasoning, 19*, 249–265.
Elqayam, S., Thompson, V. A., Wilkinson, M. R., Evans, J. St. B. T., & Over, D. E. (2015). Deontic introduction: A theory of inference from is to ought. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1516–1532.
Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Brighton UK: Erlbaum.
Evans, J. St. B. T. (2006). The heuristic–analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*, 378–395.
Evans, J. St. B. T. (2007a). *Hypothetical thinking: Dual processes in reasoning and judgement.* Hove UK: Psychology Press.
Evans, J. St. B. T. (2007b). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*, 321–339.
Evans, J. St. B. T. (2010). *Thinking twice: Two minds in one brain.* Oxford: Oxford University Press.
Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*, 295–306.
Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology, 63*, 892–909.
Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 321–355.
Evans, J. St. B. T., Neilens, H., Handley, S. J., & Over, D. E. (2008). When can we say "if"? *Cognition, 108*, 100–116.
Evans, J. St. B. T., & Over, D. E. (2004). *If.* Oxford: Oxford University Press.
Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241.
Fairchild, M. D. (2013). *Color appearance models.* Hoboken NJ: Wiley.
Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39*, 1327–1343.
Fillenbaum, S. (1976). Inducements: on phrasing and logic of conditional promises, threats and warnings. *Psychological Research, 38*, 231–250.
Fillenbaum, S. (1986). The use of conditionals in inducements and deterrents. In E. C. Traugott, A. T. Muelen, J. S. Reilly, & C. A. Ferguson (Eds.). *On Conditionals*. Cambridge: Cambridge University Press.
Fox, J. (2008). *Applied regression analysis and generalized linear models.* Los Angeles: Sage.
Fugard, A., Pfeifer, N., Mayerhofer, B., & Kleiter, G. (2011). How people interpret conditionals: Shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 635–648.
Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review, 29*, 249–282.
Gilio, A., & Over, D. E. (2012). The psychology of inferring conditionals from disjunctions: A probabilistic study. *Journal of Mathematical Psychology, 56*, 118–131.
Grice, H. P. (1989). Indicative conditionals. *Studies in the way of words* (pp. 58–85). Cambridge MA: Harvard University Press.
Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review, 114*, 704–732.
Hall, S., Ali, N., Chater, N., & Oaksford, M. (2016). Discounting and augmentation in causal conditional reasoning: causal models or shallow encoding? *PLoS ONE, 11*, e0167741. http://dx.doi.org/10.1371/journal.pone.0167741.
Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 37, 28–43.
Jackson, F. (1979). On assertion and indicative conditionals. *Philosophical Review, 88*, 565–589.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.
Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 19*, 646–678.
Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences, 19*, 201–214.
Juhos, C., Quelhas, A. C., & Johnson-Laird, P. N. (2012). Temporal and spatial relations in sentential reasoning. *Cognition, 122*, 393–404.
Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107*, 852–884.
Klauer, K. C., & Singmann, H. (2013). Does logic feel good? Testing for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning Memory and Cognition, 39*, 1265–1273.
Kratzer, A. (1986). Conditionals. In A. M. Farley, P. Farley, & K. E. McCollough (Eds.). *Papers from the parasession on pragmatics and grammatical theory* (pp. 115–135). Chicago: Chicago Linguistics Society.
Krzyżanowska, K. H., Collins, P. J., & Hahn, U. (2017). Between a conditional's antecedent and its consequent: Discourse coherence vs. probabilistic relevance. *Cognition, 164*, 199–205.
Krzyżanowska, K. H., Wenmackers, S., & Douven, I. (2014). Rethinking Gibbard's riverboat argument. *Studia Logica, 102*, 771–792.
Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review, 85*, 297–315.
Manktelow, K. I., Over, D. E., & Elqayam, S. (Eds.). (2011). *The science of reason: A festschrift in honour of Jonathan St. B. T. Evans*. Hove UK: Psychology Press.
Marr, D. (1982). *Vision.* Cambridge MA: MIT Press.
Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–74.
Mill, J. S. (1843/1872). *A system of logic* (8th ed.). London: Longmans, Green, Reader, and Dyer.
Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608–631.
Oaksford, M., & Chater, N. (2003). Conditional probability and the cognitive science of conditional reasoning. *Mind & Language, 18*, 359–379.
Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning.* Oxford: Oxford University Press.
Oaksford, M., & Chater, N. (2012). Dual processes, probabilities, and cognitive architecture. *Mind & Society, 11*, 15–26.
Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning, 19*, 346–379.
Oaksford, M., & Chater, N. (2014). Probabilistic single function dual process theory and logic programming as approaches to non-monotonicity in human vs. *artificial reasoning. Thinking & Reasoning, 20*, 269–295.
Oaksford, M., & Stenning, K. (1992). Reasoning with conditional containing negated constituents. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 835–854.
Oberauer, K., Weidenfeld, A., & Fischer, K. (2007). What makes us believe a conditional? The roles of covariation and causality. *Thinking & Reasoning, 13*, 340–369.
Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 688–693.
Ohm, E., & Thompson, V. A. (2006). Conditional probability and pragmatic conditionals: Dissociating truth and effectiveness. *Thinking & Reasoning, 12*, 257–280.
Over, D. E. (2011). New paradigm psychology of reasoning. *Thinking & Reasoning, 15*, 431–438.
Over, D. E., & Cruz, N. (2017). Probabilistic accounts of conditional reasoning. In L. J. Ball, & V. A. Thompson (Eds.). *International handbook of thinking and reasoning* (pp. 434–450). Oxon, UK: Routledge.
Over, D. E., & Baratgin, J. (2017). The "defective" truth table: Its past, present, and future. In E. Lucas, N. Galbraith, & D. E. Over (Eds.). *The thinking mind: A Festschrift for Ken Manktelow* (pp. 15–28). Hove UK: Psychology Press.
Over, D. E., Douven, I., & Verbrugge, S. (2013). Scope ambiguities and conditionals. *Thinking & Reasoning, 19*, 284–307.
Over, D. E., & Evans, J. St. B. T. (2003). The probability of conditionals: The psychological evidence. *Mind & Language, 18*, 340–358.
Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology, 54*, 62–97.
Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both intuitive and neglected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 544–554.
Pfeifer, N., & Kleiter, G. D. (2010). The conditional in mental probability logic. In M. Oaksford, & N. Chater (Eds.). *Cognition and conditionals: Probability and logic in human thinking* (pp. 153–173). Oxford: Oxford University Press.
Politzer, G., Over, D. E., & Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning, 16*, 172–197.
Ramsey, F. P. (1929/1990). General propositions and causality. In Mellor, D. H. (Ed.), *Philosophical papers* (pp. 145–163). Cambridge: Cambridge University Press.
Récanati, F. (2000). *Oratio obliqua: An essay on metarepresentation.* Cambridge MA: MIT Press.
Schroyens, W. (2010). A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions. *European Journal of Cognitive Psychology, 22*, 897–921.
Simon, H. A. (1982). *Models of bounded rationality.* Cambridge MA: MIT Press.
Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*, 559–569.
Singmann, H., Klauer, K. C., & Beller, S. (2016). Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology, 88*, 61–87.
Singmann, H., Klauer, K. C., & Over, D. E. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*. http://dx.doi.org/10.3389/fpsyg.2014.0031.
Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition, 150*, 26–32.
Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.). *Studies in logical theory* (pp. 98–112). Oxford: Blackwell.
Thompson, V. A., & Evans, J. St. B. T. (2012). Belief bias in informal reasoning. *Thinking & Reasoning, 18*, 278–310.
Thompson, V. A., & Johnson, S. J. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning, 20*, 215–244.
Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*, 107–140.
Thompson, V. A., Prowse Turner, J. A., Pennycook, G., Ball, L., Barak, H., Yael, O., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency in the monitoring and control of reasoning. *Cognition, 128*, 237–251.
van Wijnbergen-Huitink, J., Elqayam, S., & Over, D. E. (2015). The probability of iterated conditionals. *Cognitive Science, 39*, 788–803.
Vidal, M., & Baratgin, J. (2017). A psychological study of unconnected conditionals. *Journal of Cognitive Psychology, 29*, 769–781.
Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.). *New horizons in psychology* (pp. 106–137). Harmondsworth: Penguin.
Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika, 92*, 937–950.
Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology, 44*, 41–61.